

**IMPLEMENTASI MODEL SARIMAX UNTUK PERAMALAN VOLUME
PENUMPANG KERETA API DI PULAU JAWA**

SKRIPSI

Oleh:
ISNI ROHAIDAH
NIM. 19650073



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

**IMPLEMENTASI MODEL SARIMAX UNTUK PERAMALAN VOLUME
PENUMPANG KERETA API DI PULAU JAWA**

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
ISNI ROHAIDAH
NIM. 19650073

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026**

HALAMAN PERSETUJUAN

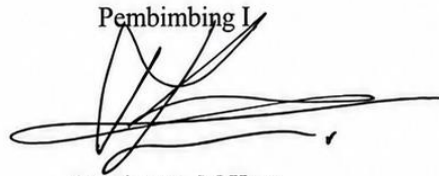
**IMPLEMENTASI MODEL SARIMAX UNTUK PERAMALAN VOLUME
PENUMPANG KERETA API DI PULAU JAWA**

SKRIPSI

Oleh :
ISNI ROHAIDAH
NIM. 19650073

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 19 Juni 2026

Pembimbing I,



Supriyono, M.Kom
NIP. 19841010 201903 1 012

Pembimbing II,



Shoffin Nahwa Utama, M.T
NIP. 19860703 202012 1 003

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M.Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

IMPLEMENTASI MODEL SARIMAX UNTUK PERAMALAN VOLUME PENUMPANG KERETA API DI PULAU JAWA

SKRIPSI

Oleh :
ISNI ROHAIDAH
NIM. 19650073

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 22 Juni 2026

Susunan Dewan Penguji

Ketua Penguji	: <u>Prof. Dr. Muhammad Faisal, M.T</u> NIP. 19740510 200501 1 007
Anggota Penguji I	: <u>Ashri Shabrina Afrah, M.T</u> NIP. 19900430 202012 2 003
Anggota Penguji II	: <u>Supriyono, M.Kom</u> NIP. 19841010 201903 1 012
Anggota Penguji III	: <u>Shoffin Nahwa Utama, M.T</u> NIP. 19860703 202012 1 003

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Isn Rohaidah

NIM : 19650073

Fakultas / Jurusan : Sains dan Teknologi / Teknik Informatika

Judul Skripsi : Implementasi Model SARIMAX untuk Peramalan Volume
Penumpang Kereta Api di Pulau Jawa

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 26 Juni 2026

Yang membuat pernyataan,



Isn Rohaidah
NIM. 19650073

MOTTO

don't compare your day 1 to someone's day 792

HALAMAN PERSEMBAHAN

Alhamdulillah rabbil 'alamin, segala puji syukur kehadiran Allah SWT yang telah memberikan jalan terbaik dalam menyelesaikan studi S1 di Universitas Islam Negeri Maulana Malik Ibrahim Malang. Sholawat serta salam selalu tercurahkan kepada Nabi Muhammad SAW yang telah membawa risalah kebenaran sebagai petunjuk hidup kita di dunia ini.

Skripsi ini penulis persembahkan untuk diri sendiri yang sudah berusaha semaksimal mungkin dan tidak menyerah dalam menyelesaikan skripsi ini. Kepada kedua orang tua yang selalu mendoakan dan memberikan dukungan. Kepada kedua kakak penulis yang selalu memberikan dukungan. Kepada seluruh dosen Jurusan Teknik Informatika UIN Maulana Malik Ibrahim Malang yang telah membimbing dan mengarahkan pengerjaan skripsi.

Kepada sahabat, teman dan orang-orang yang penulis sayangi, yang tidak bisa disebutkan satu persatu serta semua pihak yang ikut membantu dalam menyelesaikan skripsi ini.

KATA PENGANTAR

Puji syukur kepada Allah SWT yang senantiasa memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Model SARIMAX untuk Peramalan Volume Penumpang Kereta Api Di Pulau Jawa” dengan baik. Tujuan peneliti mengerjakan skripsi ini guna memenuhi syarat kelulusan dan mendapatkan gelar sarjana komputer (S.Kom) di program studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Selama penyusunan skripsi ini, penulis banyak mendapatkan dukungan, bantuan, bimbingan, serta arahan dari berbagai pihak. Oleh karena itu, dengan kerendahan hati penulis menyampaikan terima kasih kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. H. Agus Mulyono, M.Si, selaku dekan Fakultas Sains dan teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom selaku Ketua Prodi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Supriyono, M.Kom selaku Dosen Pembimbing I yang telah memberikan arahan, bimbingan, dan dukungan selama penulisan skripsi ini sehingga dapat menyelesaikan skripsi ini dengan baik.
5. Shoffin Nahwa Utama, M.T selaku Dosen Pembimbing II yang sudah membantu memberikan arahan, bimbingan, dan dukungan selama dan

dukungan selama penulisan skripsi ini sehingga dapat menyelesaikan skripsi ini dengan baik.

6. Prof. Dr. Muhammad Faisal, M.T selaku Dosen Penguji I dan Asri Shabrina Afrah, M.T selaku Dosen Penguji II yang sudah memberikan saran terbaiknya dan sudah meluangkan waktunya untuk menguji penulis dalam skripsi ini.
7. Bapak dan Ibu dosen serta staf Program Studi Teknik Informatika yang telah memberikan ilmu dan membantu proses perkuliahan sampai penyusunan skripsi ini selesai.
8. Ibu, Bapak dan kedua kakak penulis yang telah memberikan banyak dukungan, doa, dan pengorbanan kepada penulis selama proses belajar hingga akhir pengerjaan skripsi ini.
9. Teman-teman Angkatan 2019 “ALIEN” Teknik Informatika yang telah mendukung penulis dalam penyusunan skripsi ini.
10. Keluarga, sahabat, teman-teman yang telah memberikan semangat, perhatian, doa selama masa perkuliahan, sehingga skripsi ini dapat terselesaikan

Malang, 19 Juni 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	Error! Bookmark not defined.
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
ABSTRAK	xiv
ABSTRACT	xv
مستخلص البحث.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	6
1.3 Tujuan Penelitian	7
1.4 Batasan Masalah.....	7
1.5 Manfaat Penelitian	7
BAB II STUDI PUSTAKA	8
2.1 Penelitian Terkait	8
2.2 <i>Time Series</i>	12
2.2.1 Komponen Non-Musiman.....	13
2.2.2 Komponen Musiman	13
2.3 Efek Kalender dan <i>Moving Holiday</i>	14
2.4 ARIMA (<i>AutoRegressive Integrated Moving Average</i>).....	15
2.5 SARIMA (<i>Seasonal Autoregressive Integrated Moving Average</i>).....	16
2.6 SARIMAX (<i>Seasonal Autoregressive Integrated Moving Average with Exogenous Variables</i>)	20
2.6.3 <i>Preprocessing</i>	24
2.6.3.1 <i>Data Cleaning</i>	25
2.6.3.2 <i>Data Formatting</i>	25
2.6.3.3 Integrasi Variabel Eksogen	26
2.6.4 Pengujian Diagnostik Residual	26
2.6.5 Evaluasi Akurasi Peramalan	27
BAB III METODOLOGI PENELITIAN	29
3.1 Alur Penelitian	29
3.2 Pengumpulan Data	30
3.3 Desain Sistem.....	30
3.4 Metode SARIMAX	33
3.4.1 Input Data Aktual dan Data Eksogen.....	34
3.4.2 Pra-Pemrosesan Data	35
3.4.2.1 <i>Data Cleaning</i>	35
3.4.2.2 <i>Date Formatting</i>	36

3.4.2.3 Integrasi Variabel Eksogen	36
3.4.3 Uji Stasioneritas	37
3.4.4 Differencing	37
3.4.4.1 Non-Seasonal Differencing (d)	38
3.4.4.2 Seasonal Differencing (D).....	38
3.4.4.3 Penentuan Orde Diferensiasi.....	39
3.4.5 Identifikasi Model SARIMAX.....	39
3.4.6 Menentukan Model	40
3.4.7 Estimasi dan Uji Parameter	40
3.4.8 Uji Diagnostik	40
3.4.9 Skenario Pengujian	41
BAB IV HASIL DAN PEMBAHASAN	44
4.1 Uji Stasioneritas	44
4.1.1 Non-Seasonal Differencing (d)	45
4.1.2 Seasonal Differencing	46
4.2 Identifikasi Model	47
4.3 Estimasi dan Uji Signifikansi Parameter.....	50
4.4 Uji Kesesuaian Model	52
4.5 Skenario Pengujian.....	53
4.6 Pembahasan.....	67
4.7 Integrasi Islam	71
BAB V KESIMPULAN DAN SARAN	73
5.1 Kesimpulan	73
5.2 Saran.....	73
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 2.1 Arsitektur SARIMA	19
Gambar 2.2 Arsitektur SARIMAX	22
Gambar 3.1 Alur Penelitian.....	29
Gambar 3.2 Desain Sistem.....	31
Gambar 3.3 Flowchart Proses Metode SARIMAX	33
Gambar 4.1 Plot ACF dan PACF Setelah Differencing.....	48
Gambar 4.2 Grafik Hasil Peramalan Rasio 90:10	57
Gambar 4.3 Grafik Hasil Peramalan Rasio 80:20	61
Gambar 4.4 Grafik Hasil Peramalan Rasio 70:30	65

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	10
Tabel 2.2 Perbandingan SARIMA dan SARIMAX.....	23
Tabel 2.3 Kriteria Interpretasi Nilai MAPE	27
Tabel 3.1 Jumlah Penumpang KAI Tahun 2020-2025.....	35
Tabel 3.2 Matriks Frekuensi Hari Libur (Tahun 2020-2025)	35
Tabel 3.3 Data Transformasi Jumlah Penumpang KAI	36
Tabel 3.4 Integrasi Variabel Eksogen	37
Tabel 3.5 Skenario Pengujian	42
Tabel 4.1 Uji Stasioner.....	45
Tabel 4.2 Uji Stasioner Differencing Pertama	45
Tabel 4.3 Uji Stasioneritas Differencing Kedua	46
Tabel 4.4 Uji Stasioner Seasonal Differencing	46
Tabel 4.5 Model Sementara yang Digunakan	49
Tabel 4.6 Hasil Uji Signifikansi Parameter Pada Setiap Model	50
Tabel 4.7 Hasil Uji White Noise	52
Tabel 4.8 Pengujian MAPE Pertama	53
Tabel 4.9 Pengujian RMSE Pertama.....	55
Tabel 4.10 Pengujian MAPE Kedua	58
Tabel 4.11 Pengujian RMSE Kedua	60
Tabel 4.12 Pengujian MAPE Ketiga.....	62
Tabel 4.13 Pengujian RMSE Ketiga	64
Tabel 4.14 Hasil Skenario Pengujian	69

ABSTRAK

Rohaidah, Isni, 2026. **Implementasi Model SARIMAX untuk Peramalan Volume Penumpang Kereta Api di Pulau Jawa**. Skripsi. Jurusan Teknik Informatika, Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Supriyono, M.Kom. (II) Shoffin Nahwa Utama, M.T.

Kata Kunci: Deret Waktu, Peramalan, SARIMAX, Variabel Eksogen, Volume Penumpang Kereta Api

Volume penumpang kereta api di Pulau Jawa mengalami fluktuasi yang dipengaruhi oleh pola musiman serta faktor eksternal, seperti Hari Raya Idul Fitri, libur antar semester, dan Natal serta Tahun Baru. Kondisi tersebut menyebabkan metode peramalan yang hanya mengandalkan pola historis belum mampu menangkap pengaruh faktor eksternal secara optimal. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan model *Seasonal Autoregressive Integrated Moving Average with Exogenous Variables* (SARIMAX) dalam meramalkan volume penumpang kereta api di Pulau Jawa dengan memanfaatkan variabel eksogen berupa *dummy* hari libur. Penelitian menggunakan data bulanan volume penumpang kereta api di Pulau Jawa periode Januari 2020 hingga Desember 2025 yang diperoleh dari Badan Pusat Statistik (BPS). Tahapan penelitian meliputi *preprocessing* data, uji stasioneritas menggunakan *Augmented Dickey-Fuller* (ADF), proses *differencing*, identifikasi model menggunakan plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF), pemilihan model terbaik melalui *grid search* berdasarkan nilai *Akaike Information Criterion* (AIC), uji signifikansi parameter, uji diagnostik residual menggunakan *Ljung-Box*, serta evaluasi model menggunakan *Mean Absolute Percentage Error* (MAPE) dan *Root Mean Square Error* (RMSE). Pengujian dilakukan menggunakan tiga skenario pembagian data, yaitu 70:30, 80:20, dan 90:10. Hasil penelitian menunjukkan bahwa model terbaik adalah SARIMAX (2,1,2)(1,0,0), yang memenuhi uji signifikansi parameter dan uji diagnostik residual. Berdasarkan hasil pengujian, rasio pembagian data 80:20 menghasilkan performa terbaik dengan nilai MAPE sebesar 3,75% dan RMSE sebesar 1.887,12. Nilai MAPE tersebut menunjukkan bahwa model memiliki tingkat akurasi peramalan yang sangat baik. Dengan demikian, model SARIMAX mampu mengakomodasi pola historis, pola musiman, dan pengaruh variabel eksogen sehingga dapat digunakan sebagai alternatif dalam melakukan peramalan volume penumpang kereta api di Pulau Jawa.

ABSTRACT

Rohaidah, Isni. 2026. **Implementation of the SARIMAX Model for Forecasting Railway Passenger Volume on Java Island**. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Advisors: (I) Supriyono, M.Kom. (II) Shoffin Nahwa Utama, M.T.

Keywords: SARIMAX, Forecasting, Railway Passenger Volume, Time Series, Exogenous Variables

The volume of railway passengers on Java Island experiences fluctuations influenced by seasonal patterns and external factors, such as Eid al-Fitr, semester holidays, and Christmas and New Year holidays. These conditions make forecasting methods that rely solely on historical data unable to optimally capture the influence of external factors. Therefore, this study aims to implement the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) model to forecast railway passenger volume on Java Island by incorporating holiday dummy variables as exogenous variables. This study uses monthly railway passenger volume data on Java Island from January 2020 to December 2025 obtained from the Statistics Indonesia (BPS). The research stages include data preprocessing, stationarity testing using the Augmented Dickey-Fuller (ADF) test, differencing, model identification using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, model selection through grid search based on the Akaike Information Criterion (AIC), parameter significance testing, residual diagnostic testing using the Ljung-Box test, and model evaluation using the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). The model was evaluated using three data splitting scenarios: 70:30, 80:20, and 90:10. The results indicate that the best forecasting model is SARIMAX (2,1,2)(1,0,0), which satisfies both the parameter significance test and the residual diagnostic test. Among the evaluated data splitting scenarios, the 80:20 ratio achieved the best performance, with a MAPE of 3.75% and an RMSE of 1,887.12. The MAPE value indicates that the model provides a highly accurate forecasting performance. Therefore, the SARIMAX model is capable of capturing historical patterns, seasonal patterns, and the effects of exogenous variables, making it a suitable alternative for forecasting railway passenger volume on Java Island.

مستخلص البحث

روهيده، إيسني، 2026. تطبيق نموذج SARIMAX للتنبؤ بحجم ركاب القطار في جزيرة جاوى. البحث الجامعي. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: سوبريونو، الماجستير؛ المشرف الثاني: صفين نحو أوتاما، الماجستير.

الكلمات الرئيسية: سلاسل زمنية، تنبؤ، SARIMAX، متغير خارجي، حجم ركاب قطار

يشهد حجم ركاب القطار أو السكك الحديدية في جزيرة جاوى تقلبات تتأثر بالأنماط الموسمية والعوامل الخارجية، مثل عيد الفطر، وعطلة منتصف الفصل الدراسي، وعيد الميلاد ورأس السنة الميلادية. هذه الظروف تجعل طرق التنبؤ التي تعتمد فقط على الأنماط التاريخية غير قادرة بعد على استيعاب تأثير العوامل الخارجية بشكل أمثل. لذلك، هدف هذا البحث إلى تطبيق نموذج المتوسط المتحرك المتكامل الذاتي الموسمي مع المتغيرات الخارجية (SARIMAX) في التنبؤ بحجم ركاب السكك الحديدية في جزيرة جاوى من خلال الاستفادة من المتغيرات الخارجية على شكل متغيرات وهمية لأيام العطلة. استخدم البحث بيانات شهرية عن حجم ركاب السكك الحديدية في جزيرة جاوى للفترة من يناير 2020 حتى ديسمبر 2025 التي تم الحصول عليها من الهيئة المركزية للإحصاء (BPS). تشمل مراحل الدراسة معالجة البيانات المسبقة، اختبار الثبات باستخدام اختبار ديكي-فولر المعزز (ADF)، عملية التفاضل، تحديد النموذج باستخدام مخطط دالة الارتباط التلقائي (ACF) ودالة الارتباط التلقائي الجزئي (PACF)، اختيار أفضل نموذج من خلال البحث الشبكي استناداً إلى قيمة معيار معلومات أكايك (AIC)، اختبار دلالة المعلمات، اختبار تشخيص البواقي باستخدام اختبار ليونغ-يوكس، بالإضافة إلى تقييم النموذج باستخدام متوسط نسبة الخطأ المطلق (MAPE) وجذر متوسط مربع الخطأ (RMSE). تم إجراء الاختبارات باستخدام ثلاثة سيناريوهات لتقسيم البيانات، وهي 70:30، 80:20، و 90:10. أظهرت نتائج البحث أن أفضل نموذج هو SARIMAX (2.1.2) (1.0.0)، الذي استوفى اختبار دلالة المعلمات واختبار التشخيص للبقايا. بناءً على نتائج الاختبارات، فإن نسبة تقسيم البيانات 80:20 أعطت أفضل أداء بقيمة MAPE قدرها 3.75% و RMSE قدرها 1,887.12. تشير قيمة MAPE هذه إلى أن النموذج يتمتع بمستوى دقة تنبؤية ممتاز. وبالتالي، فإن نموذج SARIMAX قادر على استيعاب الأنماط التاريخية والأنماط الموسمية وتأثير المتغيرات الخارجية، مما يجعله خياراً بديلاً لاستخدامه في التنبؤ بحجم ركاب القطار في جزيرة جاوى.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sektor transportasi memegang peranan krusial dalam menunjang aktivitas suatu negara dan tidak dapat dipisahkan dari kehidupan masyarakat. Sebagai negara kepulauan, Indonesia memiliki ketergantungan yang tinggi terhadap layanan transportasi umum sebagai sarana mobilitas antarwilayah dan antarkota. Transportasi umum yang tersedia sangat beragam mulai dari transportasi darat, laut, hingga udara. Berdasarkan data Badan Pusat Statistik, pada bulan Januari-April 2025 data penumpang transportasi darat khususnya kereta api mencapai 171,1 juta orang, data penumpang transportasi udara mencapai 19,3 juta orang, sedangkan data penumpang transportasi laut mencapai 10,2 juta orang (Badan Pusat Statistik, 2025a). Merujuk pada data tersebut dapat disimpulkan bahwa masyarakat lebih memilih moda transportasi darat khususnya kereta api dibanding transportasi umum lainnya.

Transportasi kereta api merupakan salah satu infrastruktur krusial yang melayani mobilisasi ribuan penduduk setiap harinya karena memiliki keunggulan yaitu ketepatan waktu dan kapasitas penumpang yang besar. Ketepatan dan waktu tempuh yang lebih singkat dibanding transportasi darat lainnya menjadi keunggulan karena kereta api memiliki jalur tersendiri sehingga terhindar dari kemacetan lalu lintas. Hal tersebut didasarkan pada landasan hukum Undang-Undang Nomor 23

Tahun 2007 tentang Perkeretaapian, ditegaskan bahwa jalur kereta api sebagaimana dimaksud dalam Pasal 34 ayat (1) hanya digunakan untuk pengoperasian kereta api.

Meskipun kereta api menjadi transportasi massal yang paling banyak digunakan secara nasional, distribusi volume penumpang di Indonesia menunjukkan adanya perbedaan yang signifikan. Pengguna layanan transportasi kereta api didominasi oleh penumpang yang berada di Pulau Jawa yakni melebihi 90% total penumpang nasional (Badan Pusat Statistik, 2023). Hal ini terjadi karena Pulau Jawa merupakan pusat populasi nasional. Mengacu pada data kependudukan yang dirilis oleh Badan Pusat Statistik (BPS), Pulau Jawa dihuni oleh 55.5% dari total populasi nasional (2025b). Dengan proporsi populasi yang dominan tersebut, penelitian ini difokuskan pada penumpang kereta api di Pulau Jawa, mengingat jumlahnya menunjukkan perbedaan yang signifikan dibandingkan dengan volume penumpang di Pulau Sumatra maupun Sulawesi.

Besarnya skala mobilitas di Pulau Jawa menyebabkan jumlah penumpang selalu naik-turun secara tidak menentu. Kenaikan jumlah penumpang umumnya muncul pada periode-periode tertentu, khususnya pada bulan yang bertepatan dengan hari libur atau cuti bersama, sedangkan pada bulan tanpa momen liburan cenderung terjadi penurunan volume penumpang (Mendila et al., 2023). Selain kedua momen tersebut, lonjakan volume penumpang yang stabil juga terjadi pada periode Juni hingga Juli yang bertepatan dengan masa libur antar-semester atau kenaikan kelas. Lonjakan signifikan biasanya terlihat ketika masyarakat melakukan perjalanan mudik menjelang Hari Raya Idul Fitri dan akhir tahun pada bulan Desember (Hayati et al., 2021). Lonjakan penumpang yang terjadi pada saat hari

raya, libur antar semester dan akhir tahun memiliki karakteristik yang berbeda. Libur akhir tahun memiliki tanggal yang sama setiap tahun, namun hari raya keagamaan seperti Idul Fitri selalu bergeser setiap tahunnya (Hayati et al., 2021). Sebagai langkah antisipatif terhadap potensi peningkatan jumlah penumpang kereta api, diperlukan suatu metode peramalan yang mampu memperkirakan volume penumpang pada beberapa periode mendatang.

Peramalan jumlah penumpang kereta api tidak cukup hanya menggunakan model deret waktu (*time series*) biasa. Hal ini dikarenakan adanya faktor eksternal yang signifikan memengaruhi fluktuasi data, yaitu variabel hari libur yang polanya tidak selalu tetap setiap tahunnya. Atas dasar pertimbangan tersebut, penelitian ini menerapkan metode SARIMAX (*Seasonal Autoregressive Integrated Moving Average with Exogenous Factors*). Kelebihan model ini terletak pada kapasitasnya dalam menangkap pola musiman sekaligus memasukkan variabel eksternal, yaitu hari libur ke dalam model. Dengan menambahkan variabel *dummy* hari libur sebagai variabel eksogen, model SARIMAX diharapkan memberikan estimasi yang lebih peka terhadap lonjakan penumpang yang bersifat sementara namun secara berulang (Pakaya et al., 2025).

Pentingnya penggunaan model yang adaptif ini diperkuat oleh penelitian dari Yunianto & Muklason (2025), yang menyatakan bahwa peramalan permintaan tiket berperan krusial dalam mendukung perencanaan operasional PT KAI. Dalam penelitiannya, mereka menyarankan pengembangan model yang mempertimbangkan variabel eksternal seperti hari libur untuk meningkatkan akurasi prediksi. Dengan mengintegrasikan variabel tersebut, model SARIMAX

diharapkan dapat menangkap efek lonjakan penumpang secara lebih presisi, sehingga risiko penumpukan penumpang pada periode puncak dapat diminimalkan.

Upaya melakukan peramalan yang akurat ini pada dasarnya merupakan bentuk implementasi dari nilai-nilai perencanaan dalam Islam. Sebagaimana diperintahkan dalam Al-Qur'an Surah Al-Hasyr ayat 18:

يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَلْتَنظُرْ نَفْسٌ مَّا قَدَّمَتْ لِغَدٍ وَاتَّقُوا اللَّهَ ۚ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ

“Hai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah setiap diri memperhatikan apa yang telah diperbuatnya untuk hari esok (masa depan); dan bertakwalah kepada Allah, sesungguhnya Allah Maha Mengetahui apa yang kamu kerjakan.” (Q.S. Al-Hasyr: 18).

Ayat ini menekankan pentingnya melakukan persiapan dan perhitungan matang di masa kini untuk menghadapi masa depan. Dalam konteks transportasi, menganalisis data masa lalu melalui metode statistik adalah bentuk ikhtiar untuk memprediksi kebutuhan di masa depan. Hal ini bertujuan agar pelayanan publik menjadi lebih efisien, mencegah kemudharatan akibat ketidaksiapan sarana, dan memastikan sumber daya dikelola dengan penuh tanggung jawab demi kemaslahatan masyarakat.

Perintah untuk memperhatikan masa depan tersebut selaras dengan upaya ilmiah dalam meminimalkan ketidakpastian melalui perhitungan yang akurat. Beberapa penelitian terbaru menunjukkan bahwa tingkat akurasi peramalan cenderung lebih tinggi ketika model tidak hanya bertumpu pada pola data masa lalu, tetapi juga mempertimbangkan variabel eksternal yang relevan. Yunianto & Muklason (2025) menjelaskan bahwa estimasi permintaan tiket kereta api di Indonesia membutuhkan model yang mampu menyesuaikan diri terhadap faktor di

luar pola historis, seperti periode hari libur, agar perencanaan operasional dapat berjalan lebih efektif dan efisien. Sejalan dengan itu, Kurnia Hidayat & Wijayanto (2025) menemukan bahwa pendekatan SARIMAX memberikan hasil yang lebih unggul dibandingkan SARIMA konvensional dalam mengolah data yang memiliki pola musiman sekaligus dipengaruhi variabel eksternal. Dalam penelitian tersebut, keterlibatan variabel eksogen berhasil menurunkan tingkat kesalahan prediksi hingga mencapai nilai MAPE 5,78%, yang menunjukkan performa model yang sangat baik. Temuan ini memperlihatkan bahwa memasukkan faktor luar ke dalam model deret waktu dapat meningkatkan ketelitian estimasi secara signifikan. Berdasarkan keberhasilan penggunaan SARIMAX dalam berbagai studi dengan karakteristik data yang dipengaruhi variabel eksternal, penelitian ini menerapkan pendekatan serupa untuk memodelkan jumlah penumpang kereta api di Pulau Jawa dengan memasukkan variabel hari libur sebagai faktor eksogen dalam proses peramalan.

Meskipun berbagai penelitian telah membahas peramalan pada sektor transportasi, masih terdapat beberapa keterbatasan yang dapat ditemukan. Sebagian penelitian menggunakan model SARIMA yang hanya mengandalkan pola historis data sehingga belum mempertimbangkan pengaruh faktor eksternal. Di sisi lain, penelitian yang telah menerapkan SARIMAX umumnya hanya menggunakan satu jenis variabel eksogen atau diterapkan pada objek penelitian yang berbeda. Selain itu, belum banyak penelitian yang secara khusus mengintegrasikan beberapa efek kalender sekaligus, yaitu Hari Raya Idul Fitri, libur antar semester, serta Natal dan Tahun Baru. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan

tersebut melalui penerapan model SARIMAX dengan kombinasi beberapa variabel kalender, yaitu Hari Raya Idul Fitri, libur antar semester, serta Natal dan Tahun Baru, yang relevan terhadap pola pergerakan penumpang kereta api.

Berdasarkan keseluruhan uraian tersebut, terlihat bahwa fluktuasi jumlah penumpang kereta api di Pulau Jawa dipengaruhi oleh tren, musiman, serta pengaruh kalender berupa Hari Raya Idul Fitri, libur antar semester serta Nataru. Implementasi model SARIMAX menjadi pendekatan yang relevan untuk menangani kompleksitas tersebut secara bersamaan. Kesenjangan penelitian yang telah diidentifikasi menunjukkan perlunya kajian empiris berbasis data terbaru dan variabel kalender spesifik Indonesia. Atas dasar itu, penelitian ini difokuskan pada penerapan sekaligus pengujian model SARIMAX untuk memperkirakan jumlah penumpang kereta api di Pulau Jawa dengan mempertimbangkan faktor eksogen hari libur. Hasil penelitian diharapkan memberikan kontribusi dalam pengembangan model deret waktu serta mendukung perencanaan operasional transportasi nasional secara lebih efektif.

1.2 Identifikasi Masalah

Permasalahan utama yang diangkat dalam penelitian ini adalah bagaimana tingkat keakuratan model SARIMAX dalam memperkirakan jumlah penumpang KAI dengan memasukkan variabel hari libur sebagai komponen eksogen dalam model?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk mengevaluasi serta menganalisis tingkat ketepatan model SARIMAX dalam memperkirakan jumlah penumpang KAI dengan mempertimbangkan variabel hari libur sebagai faktor pendukung dalam peramalan.

1.4 Batasan Masalah

1. Penelitian ini memanfaatkan data bulanan jumlah penumpang KAI dalam rentang waktu 2020 hingga 2025 sebagai dasar analisis.
2. Penelitian difokuskan pada volume penumpang kereta api yang beroperasi di wilayah Pulau Jawa.
3. Variabel eksogen yang digunakan dibatasi pada hari libur yang identik dengan peningkatan jumlah penumpang, khususnya periode Idul Fitri, libur antar semester, dan Nataru.

1.5 Manfaat Penelitian

1. Hasil penelitian ini diharapkan dapat menjadi dasar pertimbangan bagi PT KAI dalam merencanakan dan menyesuaikan kebutuhan operasional secara lebih efektif dan terarah.
2. Temuan studi ini dapat dimanfaatkan sebagai rujukan bagi peneliti berikutnya yang mengangkat topik penelitian yang sejenis.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Sebagai landasan, penelitian yang dilakukan oleh Hayati et al. (2021) berjudul “Prediksi Data Jumlah Penumpang Kereta dengan Efek Variasi Kalender pada Model SARIMA” menunjukkan bahwa model SARIMAX mampu memprediksi jumlah penumpang kereta dengan baik ketika terdapat pengaruh hari raya keagamaan yang tanggalnya berubah setiap tahun. Penelitian tersebut juga menemukan bahwa penambahan variabel *dummy* pada model menghasilkan prediksi yang lebih akurat dan lebih mampu menggambarkan kondisi sebenarnya dibandingkan model SARIMA standar yang belum mempertimbangkan efek variasi kalender.

Selanjutnya, Kurnia Hidayat & Wijayanto (2025) melakukan penelitian berjudul "Peramalan Volume Timbulan Sampah dengan Memanfaatkan Indeks Google Trends Menggunakan Metode SARIMAX". Studi ini memberikan landasan teknis mengenai prosedur pemilihan model terbaik melalui perbandingan nilai AIC. Temuan mereka menyimpulkan bahwa integrasi variabel eksternal ke dalam model SARIMAX berhasil menurunkan nilai kesalahan (*error*) secara signifikan hingga mencapai kategori peramalan yang sangat baik dibandingkan model klasik.

Selanjutnya, Utomo & Fanani (2020) melalui penelitian berjudul “Peramalan Jumlah Penumpang Kereta Api di Indonesia Menggunakan Metode *Seasonal Autoregressive Integrated Moving Average* (SARIMA)” mengkaji peramalan

jumlah penumpang kereta api dengan memanfaatkan model SARIMA. Penelitian tersebut berfokus pada pola musiman yang dipengaruhi oleh periode libur Natal dan Tahun Baru. Hasilnya menunjukkan bahwa model SARIMA sudah mampu mengenali pola musiman tahunan pada data penumpang. Namun, model ini masih memiliki keterbatasan dalam mengikuti perubahan jumlah penumpang yang bersifat tidak tetap. Temuan tersebut menjadi dasar bahwa penggunaan model SARIMAX berpotensi memberikan hasil peramalan yang lebih akurat karena mampu memasukkan pengaruh faktor eksternal ke dalam proses pemodelan.

Karakteristik data di wilayah spesifik dibahas oleh Mendila et al. (2023) dalam penelitian "Peramalan Jumlah Penumpang Kereta Api di Pulau Jawa Menggunakan Metode *Holt Winters Exponential Smoothing* dan *Fuzzy Time Series Markov Chain*". Studi ini memberikan gambaran mendalam mengenai fluktuasi dinamis penumpang kereta api di Pulau Jawa. Hasil penelitian menunjukkan bahwa data di Pulau Jawa memiliki pola musiman yang sangat kompleks, di mana perbandingan antar metode membuktikan bahwa pendekatan yang mampu mengakomodasi variasi tren dan musiman memberikan performa akurasi yang lebih stabil.

Penelitian terkait implementasi teknologi peramalan juga dilakukan oleh Khaw et al. (2025) dengan judul "Analisis *Time Series* dan Perancangan *Dashboard* untuk Memprediksi Penjualan dengan Metode Prophet dan SARIMAX". Temuan penelitian ini menegaskan bahwa SARIMAX memiliki reliabilitas yang tinggi dalam menangani faktor eksternal yang memengaruhi pola data deret waktu yang tidak beraturan. Penelitian ini memberikan wawasan mengenai pentingnya

penggunaan variabel eksogen untuk memaksimalkan presisi hasil prediksi pada sistem peramalan berbasis data.

Berikutnya, Syam (2022) menerapkan metode ARIMAX dengan mempertimbangkan efek variasi kalender untuk meramalkan data pencarian cokelat di Indonesia dan Amerika Serikat. Pada penelitian tersebut, variabel eksogen yang digunakan berupa variabel *dummy* kalender yang merepresentasikan peristiwa khusus, seperti Hari Raya Idul Fitri, Natal, dan Hari *Valentine*. Hasil penelitian menunjukkan bahwa penambahan variabel eksogen mampu meningkatkan akurasi peramalan dengan menghasilkan nilai MAPE di bawah 10%. Selain itu, model ARIMAX juga memberikan kinerja yang lebih baik dibandingkan model ARIMA maupun beberapa metode peramalan lainnya.

Terakhir, Yunianto & Muklason (2025) memberikan kontribusi melalui penelitian berjudul "Penerapan Metode *Hybrid LSTM* dan *Seasonal Decomposition* untuk Peramalan Permintaan Tiket Harian PT KAI". Hasil penelitian tersebut menunjukkan bahwa pemisahan data ke dalam komponen tren dan musiman (dekomposisi) sebelum dilakukan pemodelan dapat meningkatkan kemampuan algoritma dalam mengenali pola permintaan secara mendalam. Hal ini memperkuat alasan penggunaan metode berbasis musiman seperti SARIMAX yang secara inheren mempertimbangkan komponen-komponen tersebut dalam peramalannya.

Tabel 2.1 Penelitian Terdahulu

No.	Peneliti	Judul penelitian	Metode	Hasil
1.	Hayati et al. (2021)	Prediksi Data Jumlah Penumpang Kereta Api dengan Efek Variasi Kalender pada Model SARIMAX	SARIMAX	Variabel <i>dummy</i> hari raya efektif menangani pergeseran kalender dan meningkatkan akurasi prediksi.
2.	Kurnia Hidayat &	Peramalan Volume Timbulan Sampah	SARIMAX	Integrasi variabel eksogen (<i>Google Trends</i>)

No.	Peneliti	Judul penelitian	Metode	Hasil
	Wijayanto (2025)	dengan Memanfaatkan Indeks <i>Google Trends</i>		menghasilkan nilai MAPE 5,78% (Sangat Baik).
3.	Utomo & Fanani (2020)	Peramalan Jumlah Penumpang Kereta Api di Indonesia Menggunakan Metode <i>Seasonal Autoregressive Integrated Moving Average (SARIMA)</i>	SARIMA	Model SARIMA berhasil menangkap pola musiman Natal dan Tahun Baru dengan hasil peramalan yang stabil.
4.	Mendila et al. (2023)	Peramalan Jumlah Penumpang Kereta Api di Pulau Jawa Menggunakan Metode <i>Holt Winters Exponential Smoothing</i> dan <i>Fuzzy Time Series Markov Chain</i>	<i>Holt Winters Exponential Smoothing</i> dan <i>Fuzzy Time Series Markov Chain</i>	Data di Pulau Jawa terbukti memiliki pola musiman kompleks yang dipengaruhi oleh bulan-bulan tertentu.
5.	Khaw et al. (2025)	Analisis <i>Time Series</i> dan Perancangan <i>Dashboard</i> untuk Memprediksi Penjualan dengan Metode Prophet dan SARIMAX	Prophet dan SARIMAX	SARIMAX terbukti reliabel dalam menangani faktor eksternal dan pola data deret waktu yang tidak beraturan.
6.	Yunianto & Muklason (2025)	Penelitian Metode <i>LSTM</i> dan <i>Seasonal Decomposition</i> untuk Peramalan Permintaan Tiket Harian PT KAI	<i>LSTM</i> & <i>Seasonal Decomposition</i>	Teknik dekomposisi data (tren, musiman, sisa) secara signifikan membantu model mengenali pola permintaan tiket.
7.	Syam (2022)	<i>Application of the Autoregressive Integrated Moving Average Exogenous (ARIMAX) with Calendar Variation Effect Method for Forecasting Chocolate Data in Indonesia and the United States</i>	ARIMAX	Model ARIMAX dengan efek variasi kalender menghasilkan nilai MAPE < 10% dan memberikan akurasi peramalan yang lebih baik dibandingkan metode ARIMA dan metode lainnya.

Berdasarkan tinjauan penelitian terdahulu di atas, posisi penelitian ini adalah untuk mengisi celah (*gap*) penelitian dengan menggabungkan keunggulan metode

SARIMAX dalam menangani pola musiman sekaligus mengintegrasikan faktor eksternal berupa hari libur. Hal ini berbeda dengan penelitian sebelumnya yang mayoritas hanya menggunakan metode musiman standar atau hanya terbatas pada satu jenis hari raya saja. Dengan cakupan wilayah Pulau Jawa, penelitian ini diharapkan memberikan akurasi yang lebih presisi bagi perencanaan transportasi kereta api nasional.

2.2 Time Series

Data deret waktu merupakan salah satu jenis data yang paling banyak dijumpai dalam berbagai bidang keilmuan dan praktik penelitian kuantitatif. Ardesfira et al. (2022) mendefinisikan bahwa *time series* merupakan serangkaian nilai-nilai variabel yang disusun berdasarkan waktu. Pemahaman yang mendasar terhadap sifat dan karakteristik data deret waktu merupakan prasyarat yang tidak dapat diabaikan sebelum peneliti menentukan pendekatan pemodelan yang akan digunakan. Tanpa pemahaman ini, pemilihan model yang keliru dapat menghasilkan prediksi dan kesimpulan yang tidak akurat.

Setiap deret waktu pada umumnya tersusun atas beberapa komponen yang membentuk pola keseluruhan data secara bersama-sama. Heru Widiyanto & Mayasari (2023) menyebutkan bahwa terdapat beberapa jenis pola pada deret waktu yaitu, tren (kecenderungan jangka panjang), musiman (fluktuasi periodik tetap), siklus (naik-turun jangka menengah tanpa periode tetap), dan variasi acak (fluktuasi tidak terduga). Dalam praktiknya, terdapat dua pendekatan dekomposisi untuk memisahkan komponen tersebut, yaitu model aditif untuk pola musiman yang konstan, dan model multiplikatif jika pola musiman membesar seiring pertumbuhan

tren (Hyndman & Anthanasopoulos George, 2021). Data volume penumpang kereta api di Pulau Jawa memiliki karakteristik kompleks karena menggabungkan tren pertumbuhan jangka panjang dengan pola musiman yang kuat, terutama pada periode hari raya. Kondisi ini membuat pemilihan metode dekomposisi dan pemodelan yang tepat menjadi langkah yang sangat menentukan kualitas hasil peramalan.

2.2.1 Komponen Non-Musiman

Komponen non-musiman merupakan pola perubahan data yang tidak terjadi secara berulang pada periode waktu tertentu. Komponen ini menggambarkan variasi dasar dalam data yang dapat berupa tren jangka panjang maupun fluktuasi yang tidak dipengaruhi oleh siklus musiman. Komponen non-musiman terdiri atas *Autoregressive* (AR), *Differencing* (I), dan *Moving Average* (MA). Dalam pemodelan deret waktu, Mirdan Faris et al. (2024) menjelaskan bahwa parameter (p,d,q) pada model ARIMA digunakan untuk merepresentasikan komponen non-musiman, sedangkan parameter (P,D,Q) digunakan untuk menggambarkan komponen musiman. Dengan demikian, komponen non-musiman berperan dalam menangkap hubungan antarobservasi berdasarkan pola historis data tanpa mempertimbangkan adanya pengulangan yang terjadi secara periodik.

2.2.2 Komponen Musiman

Komponen musiman merupakan pola perubahan data yang muncul secara berulang dalam selang waktu tertentu, seperti harian, mingguan, bulanan, atau tahunan. Pola ini umumnya dipengaruhi oleh peristiwa yang terjadi secara rutin

sehingga menyebabkan kenaikan maupun penurunan nilai data pada periode yang relatif sama setiap siklusnya. Dalam analisis deret waktu, komponen musiman terdiri atas *Seasonal Autoregressive* (SAR), *Seasonal Differencing* (SI), dan *Seasonal Moving Average* (SMA). Menurut Fungki Wahyu & Billy Hendrik (2023), komponen musiman ditunjukkan oleh pola yang terus berulang pada interval waktu yang tetap sehingga dapat dikenali dari kemunculan pola yang konsisten pada setiap periodenya. Berkaitan dengan penelitian ini, karakteristik musiman terlihat dari meningkatnya jumlah penumpang kereta api pada periode Hari Raya Idul Fitri, libur antarsemester, serta Natal dan Tahun Baru. Oleh karena itu, pola musiman tersebut perlu diperhitungkan dalam proses pemodelan agar hasil peramalan dapat menggambarkan kondisi data dengan lebih baik.

2.3 Efek Kalender dan *Moving Holiday*

Efek kalender merupakan pengaruh peristiwa tertentu dalam sistem penanggalan terhadap pola data deret waktu. Efek ini dibedakan menjadi dua jenis yaitu *fixed holiday* dan *moving holiday*. *Fixed holiday* adalah hari libur dengan tanggal yang tetap setiap tahun, seperti libur sekolah, Natal dan Tahun Baru, sehingga efeknya muncul pada periode yang sama dan relatif mudah dikenali model musiman. Sementara itu, *moving holiday* merupakan hari libur yang tanggalnya berubah setiap tahun, seperti Idul Fitri yang bergeser sekitar 10-11 hari karena mengikuti kalender Hijriah (Syam, 2022). Pergeseran ini menjadi kendala bagi SARIMA karena model hanya mengenali pola musiman tetap, sehingga lonjakan akibat Idul Fitri sering dianggap sebagai fluktuasi acak dan bukan pola yang berulang.

Untuk mengatasi masalah tersebut, model SARIMAX menggunakan variabel eksogen berupa *dummy* biner yang mengikuti waktu terjadinya hari libur. Pendekatan ini membuat model mampu mengenali pengaruh hari libur secara lebih cepat. Hayati et al. (2021) membuktikan bahwa penambahan *dummy* efek kalender pada SARIMAX mampu meningkatkan akurasi peramalan jumlah penumpang kereta api. Hasil serupa juga ditemukan oleh Syam (2022), yang menunjukkan bahwa variabel *dummy* pada ARIMAX dapat menjelaskan variasi data yang tidak dapat ditangkap oleh ARIMA biasa.

2.4 ARIMA (*AutoRegressive Integrated Moving Average*)

Model *Autoregressive Integrated Moving Average* (ARIMA) merupakan metode peramalan univariat yang digunakan untuk memodelkan data deret waktu nonstasioner tanpa pola musiman (Box et al., 2015). Model ini dibentuk dari tiga komponen utama yang dinyatakan dalam notasi ARIMA(p,d,q). Komponen pertama adalah *Autoregressive* (AR) dengan orde p , yang menjelaskan hubungan antara nilai saat ini dengan sejumlah nilai pada periode sebelumnya. Komponen kedua yaitu *Integrated* (I) dengan orde d , yang melakukan proses *differencing* sebanyak d kali untuk mengubah data nonstasioner menjadi stasioner melalui perhitungan selisih antarobservasi Hyndman & Athanasopoulos George (2021). Sementara itu, komponen ketiga adalah *Moving Average* (MA) dengan orde q , yang menggambarkan pengaruh residual atau kesalahan prediksi pada beberapa periode sebelumnya terhadap nilai pada periode saat ini.

Secara sistematis, apabila Y merupakan deret hasil *differencing* orde d , maka model ARIMA(p,d,q) diformulasikan sebagai berikut (Box et al., 2015) :

$$Y'_t = c + \phi_1 Y'_{t-1} + \phi_2 Y'_{t-2} + \dots + \phi_p Y'_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} + e_t \quad (2.1)$$

Dimana:

Y'_t	: Nilai data pada waktu t (setelah differencing)
c	: Konstanta
$\phi_1 \dots \phi_p$	: Parameter <i>Autoregressive</i> (AR)
$Y'_{t-1} \dots Y'_{t-p}$	: Data masa lalu hingga lag p
$\theta_1 \dots \theta_q$	: Parameter Moving Average (MA)
e_t	: Nilai <i>error</i> (<i>white noise</i>)

Dalam penerapannya pada data volume penumpang kereta api, model ARIMA mampu menangkap pola ketergantungan antar waktu yang ada pada data secara efektif. Namun demikian, Garini & Anbiya (2022) menegaskan bahwa ARIMA memiliki batasan mendasar karena tidak mengakomodasikan komponen musiman dalam strukturnya, sehingga apabila data memiliki pola perulangan periodik yang kuat seperti lonjakan penumpang pada periode hari raya, model ARIMA konvensional tidak dapat menangkap pola tersebut secara memadai. Keterbatasan inilah yang mendorong pengembangan model kearah SARIMA dan selanjutnya SARIMAX, sebagaimana yang menjadi inti dari penelitian ini.

2.5 SARIMA (*Seasonal Autoregressive Integrated Moving Average*)

Model Seasonal Autoregressive Integrated Moving Average (SARIMA) merupakan pengembangan dari model ARIMA yang dirancang untuk menangani data deret waktu dengan pola musiman. Jika ARIMA digunakan untuk data nonstasioner tanpa komponen musiman, maka SARIMA mampu memodelkan pola musiman secara langsung Hyndman & Anthanasopoulos George (2021). Karakteristik ini sesuai dengan data transportasi kereta api di Indonesia karena jumlah penumpang cenderung mengalami kenaikan dan penurunan yang berulang

setiap tahun, terutama pada periode Hari Raya Idul Fitri, libur sekolah, serta Natal dan Tahun Baru. Untuk menangkap pola tersebut, SARIMA menambahkan komponen musiman ke dalam struktur ARIMA sehingga dapat memodelkan pola jangka pendek dan pola musiman secara bersamaan (Box et al., 2015). Hal ini juga didukung oleh penelitian Cania et al. (2023) yang menunjukkan bahwa model SARIMA menghasilkan peramalan yang lebih akurat dengan nilai MAPE lebih rendah dibandingkan model ARIMA konvensional pada data penumpang kereta api yang memiliki pola musiman tahunan yang kuat.

Perbedaan utama antara SARIMA dan ARIMA terletak pada penambahan parameter musiman $(P, D, Q)_s$. Parameter P (*Seasonal AR*) menggambarkan pengaruh observasi pada periode musiman masa lalu terhadap nilai saat ini, sedangkan Q (*Seasonal MA*) menyatakan pengaruh error prediksi pada periode musiman sebelumnya. Untuk menstabilkan pola musiman yang tidak stasioner, dilakukan D (*Seasonal Differencing*) dengan cara mengurangi nilai observasi saat ini dengan nilai pada lag sebesar s (Box et al., 2015). Dalam penelitian ini, panjang periode musiman s ditetapkan sebesar 12, yang merepresentasikan satu siklus tahunan pada data bulanan volume penumpang kereta api. Penentuan nilai P , D , dan Q dilakukan melalui analisis plot ACF dan PACF pada lag kelipatan s , di mana lonjakan koreasi yang signifikan pada periode tersebut menjadi indikator adanya struktur musiman (Hyndman & Athanasopoulos George, 2021).

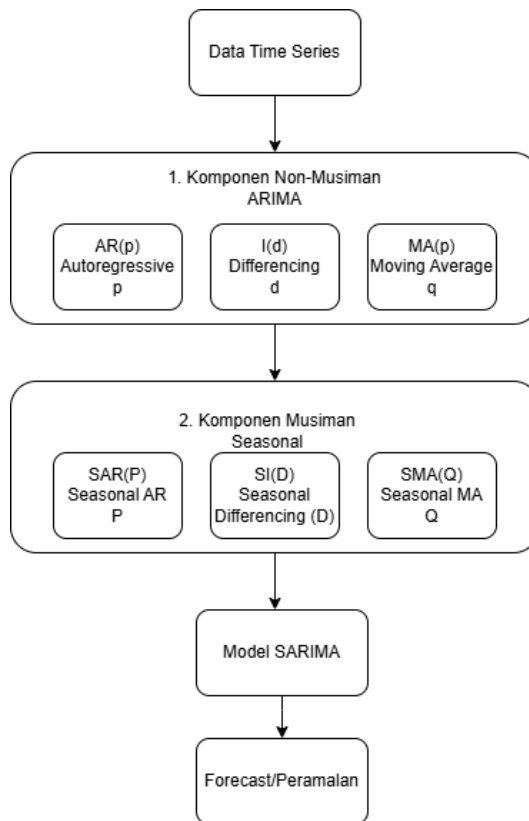
Secara formal, model SARIMA dinotasikan sebagai $SARIMA(p,d,q)(P,D,Q)_s$. persamaan umum model ini dapat dituliskan sebagai berikut (Box et al., 2015).

$$Y'_t = c + \phi_1 Y'_{t-1} + \dots + \phi_p Y'_{t-p} + \Phi_1 Y'_{t-s} + \dots + \Phi_P Y'_{t-PS} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + \Theta_1 e_{t-s} + \dots + \Theta_Q e_{t-QS} + e_t \quad (2.2)$$

Dimana:

- Y'_t : Nilai data (setelah *differencing* d dan D)
- c : Konstanta
- ϕ : Parameter AR non-musiman
- Φ : Parameter AR musiman
- θ : Parameter MA non-musiman
- Θ : Parameter MA musiman
- e_t : *White noise (error)*

Model SARIMA sangat relevan untuk data volume penumpang kereta api di Pulau Jawa karena mampu mengakomodasikan tren pertumbuhan jangka panjang dan pola musiman akibat lonjakan hari raya (Garini & Anbiya, 2022). Namun, SARIMA memiliki keterbatasan dalam menangani fenomena *moving holiday* seperti Hari Raya Idulfitri yang tanggalnya maju sekitar 10 hari setiap tahun dalam kalender Masehi (Syam, 2022). Kondisi pergeseran ini menyebabkan struktur musiman statis pada SARIMA tidak cukup akurat untuk menangkap efek hari raya yang bergerak secara dinamis antar bulan. Berdasarkan kondisi tersebut, model SARIMA perlu dikembangkan menjadi SARIMAX dengan menambahkan variabel eksogen. Variabel eksogen adalah faktor yang berasal dari luar data deret waktu, tetapi tetap memiliki pengaruh terhadap perubahan nilai data. Dengan adanya variabel ini, proses peramalan tidak hanya bergantung pada pola historis, melainkan juga mempertimbangkan faktor eksternal yang dapat memengaruhi data pada periode tertentu. Penambahan variabel eksogen bertujuan untuk menangkap pengaruh efek kalender secara lebih jelas dan terukur sehingga model SARIMAX diharapkan mampu menghasilkan tingkat akurasi peramalan yang lebih baik dibandingkan model SARIMA standar.



Gambar 2.1 Arsitektur SARIMA

Gambar 2.1 memperlihatkan arsitektur model SARIMA yang digunakan untuk memodelkan data deret waktu dengan pola musiman. Proses pemodelan dimulai dari data deret waktu sebagai masukan utama. Data tersebut kemudian diproses melalui komponen nonmusiman sebelum dilanjutkan ke komponen musiman. Setelah kedua komponen digabungkan, terbentuk model SARIMA dengan parameter $(p,d,q)(P,D,Q)$. Model yang telah diperoleh selanjutnya digunakan untuk menghasilkan nilai peramalan berdasarkan pola historis dan pola musiman yang terdapat pada data.

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

Model SARIMAX merupakan pengembangan dari model SARIMA yang mengintegrasikan variabel eksternal atau eksogen ke dalam kerangka peramalan deret waktu. Berbeda dengan model univariat yang hanya mengandalkan pola historis data, SARIMAX bersifat multivariat terbatas dengan mempertimbangkan pengaruh faktor luar yang secara signifikan mempengaruhi variabel dependen (Hyndman & Anthanasopoulos George, 2021). Secara filosofis, SARIMAX mengakui bahwa fluktuasi deret waktu di dunia nyata tidak hanya digerakkan oleh dinamika internal, tetapi juga oleh faktor eksternal sebagai penggerak tambahan (Juliarto et al., 2024). Dalam penelitian ini, variabel eksogen seperti Idulfitri, libur antar semester, Natal, dan Tahun Baru digunakan untuk menjelaskan variansi lonjakan volume penumpang kereta api di Pulau Jawa yang tidak dapat ditangkap secara akurat oleh struktur musiman statis SARIMA.

Implementasi variabel eksogen dalam ini dilakukan melalui pendekatan variabel *dummy* biner, yang bernilai pada bulan terjadinya hari raya dan 0 untuk bulan lainnya. Pendekatan ini memungkinkan model untuk mengidentifikasi lonjakan volume penumpang sebagai konsekuensi logis dari faktor eksternal, bukan sekedar fluktuasi acak (Syam, 2022). Penggunaan variabel *dummy* menjadi solusi krusial dalam menangani fenomena *moving holiday* seperti Idulfitri yang tanggalnya bergeser setiap tahun mengikuti kalender Hijriah. Sebagaimana ditegaskan oleh (Hayati et al., 2021), hari raya keagamaan merupakan efek variasi kalender yang membutuhkan akomodasi eksplisit dalam model agar pergeseran

tersebut tidak menurunkan akurasi prediksi. Dengan mengodekan kejadian hari raya mengikuti posisi aktualnya pada kalender Masehi, model dapat menangkap efek hari raya secara terukur, sehingga menghasilkan ramalan yang lebih presisi dibandingkan model tanpa variabel eksogen.

Secara sistematis, model SARIMAX(p,d,q)(P,D,Q)_s dibangun diatas persamaan SARIMA dengan penambahan komponen regresi linear dari variabel eksogen. Persamaan umum SARIMAX dapat dituliskan sebagai berikut:

$$Y'_t = c + \beta X_t + \sum_{i=1}^p \phi_i Y'_{t-i} + \sum_{j=1}^P \Phi_j Y'_{t-j_s} + \sum_{k=1}^q \theta_k e_{t-k} + \sum_{l=1}^Q \Theta_l e_{t-l_s} + e_t \quad (2.3)$$

Dimana:

- Y'_t : Nilai data setelah proses *differencing* non-musiman (d) dan musiman (D)
- c : Konstanta
- β : Koefisien regresi variabel eksogen
- X_t : Variabel eksogen (*dummy* hari raya) pada waktu t
- ϕ_p : Parameter *Autoregressive* (AR) non-musiman orde ke-p
- Φ_P : Parameter *Autoregressive* (AR) musiman orde ke-P
- θ_q : Parameter *Moving Average* (MA) non-musiman orde ke-q
- Θ_Q : Parameter *Moving Average* (MA) musiman orde ke-Q
- s : Periode musiman (misal: $s=12$ untuk data bulanan)
- e_t : Sisaan (*white noise/error*) pada waktu t

Estimasi parameter model SARIMAX dilakukan menggunakan metode *Maximum Likelihood Estimation* (MLE). Metode ini bekerja dengan mencari kombinasi nilai parameter $(\phi, \Phi, \theta, \Theta, \beta)$ yang memaksimalkan fungsi *likelihood*, yaitu probabilitas model dalam menghasilkan angka yang sesuai dengan hasil observasi (Hyndman & Athanasopoulos George, 2021). Setelah diperoleh beberapa model yang valid melalui uji diagnostik, dilakukan pemilihan model terbaik menggunakan kriteria objektif *Akaike Informaton Criterion* (AIC). Menurut Juliarto et al. (2024), AIC merupakan alat ukur yang krusial dalam menjaga keseimbangan antara akurasi model dan kesederhanaan strukturnya. Persamaan AIC dirumuskan sebagai berikut:

$$AIC = -2 \ln(\hat{L}) + 2k \quad (2.4)$$

Keterangan:

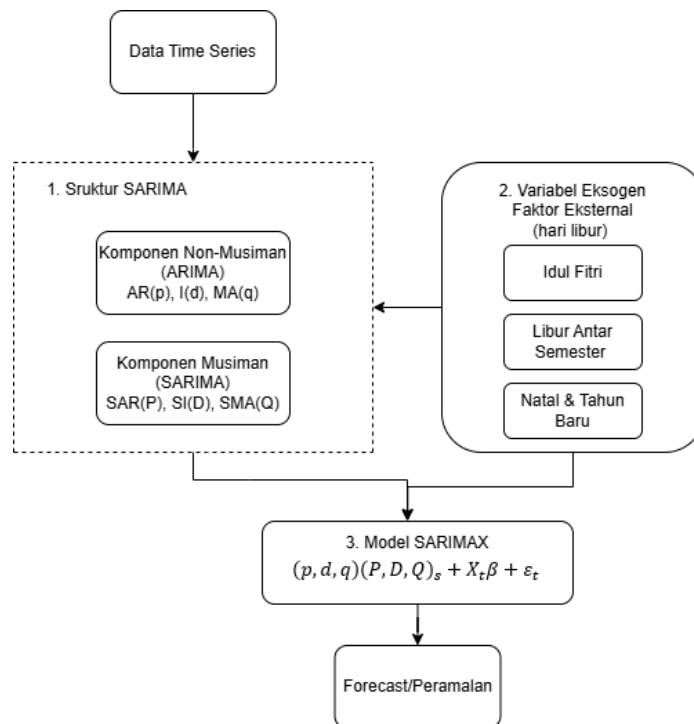
AIC : *Akaike Information Criterion* (ukuran efisiensi model)

$\ln$: Logaritma natural

$\hat{L}$: *Maximum Likelihood Function* (nilai peluang data observasi diberikan parameter model)

k : Jumlah total parameter yang diestimasi dalam model

Model yang memiliki nilai AIC paling kecil dipilih sebagai model terbaik karena mampu menjelaskan pola data secara lebih efisien dengan jumlah parameter yang tetap optimal. Pemilihan ini membantu mengurangi risiko *overfitting*, sehingga model tidak menjadi terlalu kompleks akibat penggunaan parameter yang tidak diperlukan.



Gambar 2.2 Arsitektur SARIMAX

Gambar 2.2 menunjukkan arsitektur model SARIMAX yang digunakan dalam penelitian. Proses pemodelan diawali dengan data deret waktu jumlah penumpang kereta api sebagai variabel utama. Data tersebut kemudian diproses

melalui struktur SARIMA yang terdiri atas komponen non-musiman serta komponen musiman. Kombinasi kedua komponen tersebut digunakan untuk menangkap pola historis dan pola musiman yang terdapat pada data.

Selain memanfaatkan pola internal data, model SARIMAX juga mengintegrasikan variabel eksogen berupa faktor eksternal yang diduga memengaruhi jumlah penumpang kereta api. Pada penelitian ini, variabel eksogen direpresentasikan oleh variabel *dummy* Hari Raya Idul Fitri, libur antar semester, serta Natal dan Tahun Baru. Seluruh komponen tersebut kemudian diproses dalam model SARIMAX sehingga menghasilkan nilai peramalan yang tidak hanya mempertimbangkan pola historis dan musiman, tetapi juga pengaruh faktor eksternal yang relevan terhadap pergerakan jumlah penumpang kereta api.

Tabel 2.2 Perbandingan SARIMA dan SARIMAX

Aspek	SARIMA	SARIMAX
Sumber Informasi	Hanya menggunakan data historis	Menggunakan data historis dan variabel eksogen
Komponen model	Non-musiman dan musiman	Non-musiman, musiman, dan variabel eksogen
Pengaruh faktor eksternal	Tidak dapat dimodelkan secara langsung	Dapat dimodelkan secara langsung
Kemampuan menangkap efek kalender	Terbatas pada pola historis	dapat memasukkan variabel <i>dummy</i> hari libur
Kesuaian untuk penelitian ini	Kurang optimal	Lebih sesuai karena mempertimbangkan hari libur
Akurasi peramalan	Baik jika hanya dipengaruhi pola historis	Umumnya lebih tinggi apabila terdapat faktor eksternal yang berpengaruh

SARIMA mampu memodelkan pola musiman berdasarkan data historis, tetapi belum dapat mengakomodasi pengaruh faktor eksternal secara langsung. Sebaliknya, SARIMAX merupakan pengembangan SARIMA yang memungkinkan penambahan variabel eksogen sehingga pengaruh faktor luar dapat dimasukkan ke

dalam model. oleh karena itu, SARIMAX dipilih karena lebih sesuai untuk memodelkan data yang dipengaruhi oleh pola musiman sekaligus faktor eksternal.

Sejumlah penelitian terdahulu menunjukkan bahwa model SARIMAX umumnya memberikan hasil peramalan yang lebih akurat dibandingkan SARIMA ketika data dipengaruhi oleh faktor eksternal. Salah satunya adalah penelitian Kurnia Hidayat & Wijayanto (2025) yang membandingkan kinerja kedua model dalam meramalkan volume timbunan sampah di Daerah Istimewa Yogyakarta. Hasil penelitian tersebut menunjukkan bahwa model SARIMAX(1,1,1)(1,0,0) dengan variabel eksogen berupa Indeks Produksi Industri (IBS) dan *Google Trends* menghasilkan nilai RMSE sebesar 46,7509 dan MAPE sebesar 5,7873%. Sementara itu, model SARIMA memperoleh RMSE sebesar 50,1785 dan MAPE sebesar 6,2262%. Selain memiliki nilai kesalahan yang lebih rendah, model SARIMAX juga menghasilkan nilai AIC dan BIC yang lebih kecil dibandingkan SARIMA. Temuan tersebut menunjukkan bahwa penambahan variabel eksogen membantu model menangkap pengaruh faktor di luar pola historis, sehingga performa peramalan menjadi lebih akurat.

2.6.3 Preprocessing

Sebelum digunakan dalam proses pemodelan SARIMAX, data perlu melalui beberapa tahapan preprocessing agar siap untuk dianalisis. Tahapan ini dilakukan untuk meningkatkan kualitas data, memastikan struktur deret waktunya tersusun dengan baik, serta menyiapkan variabel eksogen yang akan digunakan selama proses pemodelan.

2.6.3.1 Data Cleaning

Tahap awal dalam proses *preprocessing* adalah *data cleaning*, yang bertujuan untuk memastikan data memiliki kualitas yang baik sebelum digunakan pada tahap analisis dan pemodelan. Pada tahap ini, data diperiksa untuk menemukan berbagai permasalahan, seperti *missing value*, yang dapat memengaruhi hasil penelitian. Menurut Tsabita et al. (2025), *preprocessing* merupakan proses mengolah data mentah menjadi data yang siap digunakan dengan menghilangkan berbagai permasalahan yang berpotensi mengganggu tahapan berikutnya. Melalui proses *data cleaning*, data menjadi lebih rapi, terstruktur, dan lebih layak digunakan sebagai dasar dalam membangun model peramalan.

2.6.3.2 Data Formatting

Setelah proses *data cleaning* selesai dilakukan, data perlu melalui tahap *data formatting* agar memiliki struktur waktu yang sesuai untuk analisis deret waktu. Pada penelitian *time series*, informasi tanggal menjadi komponen yang sangat penting karena setiap observasi harus tersusun secara berurutan sesuai periode pengamatan. Lifa & Yuliansyah (2025) menjelaskan bahwa transformasi data dilakukan dengan mengubah format tanggal menjadi YYYY-MM-DD. Penyeragaman format ini membantu perangkat lunak mengenali data tanggal dengan lebih baik sehingga proses pengolahan dan analisis dapat berjalan secara optimal.

2.6.3.3 Integrasi Variabel Eksogen

Setelah data memiliki struktur waktu yang sesuai, tahap berikutnya adalah integrasi variabel eksogen. Tahap ini dilakukan dengan menggabungkan data utama dengan faktor-faktor eksternal yang diperkirakan turut memengaruhi nilai yang akan diramalkan. Mahmudi et al. (2024) menjelaskan bahwa SARIMAX merupakan pengembangan dari model SARIMA yang memungkinkan penggunaan variabel eksogen atau eksternal dalam proses pemodelan. Kehadiran variabel eksogen membuat model tidak hanya mengandalkan pola historis yang terdapat pada data, tetapi juga mempertimbangkan faktor lain yang berasal dari luar deret waktu utama.

2.6.4 Pengujian Diagnostik Residual

Sebelum digunakan untuk melakukan peramalan, model SARIMAX perlu melalui tahap diagnostik residual untuk memastikan tidak ada pola yang masih tersisa dan belum dapat dijelaskan oleh model. Salah satu pengujian yang digunakan adalah uji *White Noise (Ljung-Box)*. Pengujian ini bertujuan untuk mengetahui apakah residual masih memiliki autokorelasi. Menurut Juliarto et al. (2024), apabila nilai p-value lebih besar dari 0,05, maka hipotesis nol gagal ditolak sehingga residual dapat dinyatakan sebagai white noise. Hasil tersebut menunjukkan bahwa model telah mampu menangkap pola sistematis pada data dengan baik.

2.6.5 Evaluasi Akurasi Peramalan

Evaluasi akurasi model SARIMAX dilakukan menggunakan kombinasi dua metrik untuk memberikan gambaran performa yang komprehensif (Hyndman & Anthanasopoulos George, 2021):

1. Mean Absolute Percentage Error (MAPE)

MAPE digunakan untuk mengukur rata-rata persentase kesalahan relatif antara nilai aktual (Y_t) dan nilai hasil prediksi ($\hat{Y}_t$). Semakin kecil nilai MAPE, semakin tinggi tingkat akurasi model dalam melakukan peramalan. Menurut Juliarto et al. (2024), model dengan nilai MAPE kurang dari 10% termasuk dalam kategori sangat akurat. Rumus perhitungan MAPE ditunjukkan sebagai berikut.

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\% \quad (2.5)$$

Keterangan:

Y_t : Nilai aktual data pada periode ke- t

$\hat{Y}_t$: Nilai hasil peramalan (prediksi) pada periode ke- t

n : Jumlah total periode atau banyaknya data yang diobservasi

$|\dots|$: Nilai absolut (menghilangkan tanda negatif agar kesalahan tidak saling menghapuskan)

$\sum_{t=1}^n$: Simbol sigma yang berarti penjumlahan dari data pertama ($t = 1$) hingga data terakhir (n)

Tabel 2.3 Kriteria Interpretasi Nilai MAPE

Nilai MAPE	Tingkat Akurasi Model
< 10%	Sangat Baik / Sangat Akurat
10-20%	Baik / Akurat
20-50%	Kurang Akurat
>50%	Tidak Akurat

Sumber: Khaw et al. (2025)

MAPE digunakan untuk mengukur besarnya kesalahan hasil peramalan dalam bentuk persentase terhadap nilai aktual. Nilai MAPE yang semakin kecil menunjukkan bahwa hasil prediksi semakin mendekati data sebenarnya sehingga akurasi model menjadi lebih tinggi. Berdasarkan kriteria pada Tabel 2.2, model

yang menghasilkan nilai MAPE kurang dari 10% termasuk dalam kategori memiliki tingkat akurasi yang sangat baik atau sangat akurat.

2. *Root Mean Square Error (RMSE)*

Mengukur akar rata-rata kuadrat selisih antara nilai aktual dan prediksi. Karena bersifat kuadratik, RMSE lebih sensitif terhadap kesalahan besar atau deviasi ekstrem, sehingga efektif mendeteksi *error* pada periode lonjakan penumpang. Persamaan RMSE dirumuskan sebagai berikut.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (2.6)$$

Keterangan:

Y_t	: Nilai aktual data pada periode ke- t
$\hat{Y}_t$	: Nilai hasil peramalan (prediksi) pada periode ke- t
n	: Jumlah total periode atau banyaknya data yang diobservasi
$(Y_t - \hat{Y}_t)^2$	: Selisih antara nilai aktual dan prediksi yang dikuadratkan (memberikan bobot lebih besar pada kesalahan yang besar)
$\sqrt{}$	: Akar kuadrat untuk mengembalikan satuan nilai agar sesuai dengan satuan asli datanya.

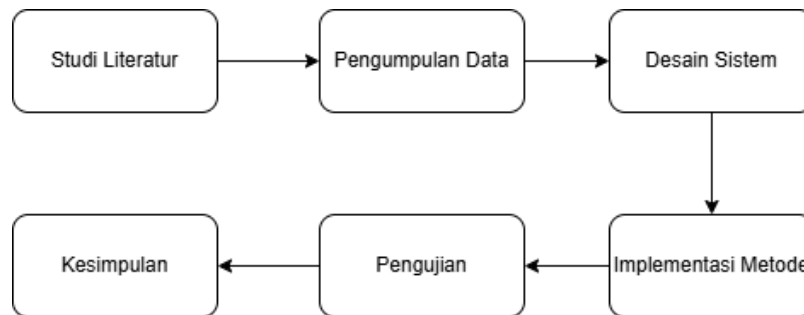
Penggunaan kedua metrik ini secara bersamaan memastikan evaluasi yang seimbang, di mana MAPE mencerminkan akurasi relatif dalam satuan persentase yang intuitif, sementara RMSE mencerminkan akurasi absolut terhadap skala data (Hyndman & Anthanasopoulos George, 2021).

BAB III

METODOLOGI PENELITIAN

3.1 Alur Penelitian

Alur studi ini memberikan gambaran mengenai step-step dari penelitian secara umum. Alur penelitian terkait implementasi metode SARIMAX dan penggunaan variabel eksogen hari libur dalam meramal jumlah penumpang kereta api di Pulau Jawa dipaparkan pada Gambar 3.1.



Gambar 3.1 Alur Penelitian

Penelitian ini dilaksanakan melalui beberapa tahapan yang disusun secara sistematis agar tujuan penelitian dapat tercapai. Proses penelitian diawali dengan pengumpulan data volume penumpang kereta api di Pulau Jawa yang bersumber dari Badan Pusat Statistik (BPS), serta data hari libur yang digunakan sebagai variabel eksogen. Setelah itu, dilakukan tahap *preprocessing* yang mencakup pembersihan data dan pembentukan variabel *dummy* hari libur. Data yang telah siap kemudian dianalisis menggunakan metode SARIMAX untuk membangun model peramalan. Pada tahap akhir, model digunakan untuk memprediksi volume penumpang kereta api, kemudian kinerjanya dievaluasi menggunakan metrik *Mean*

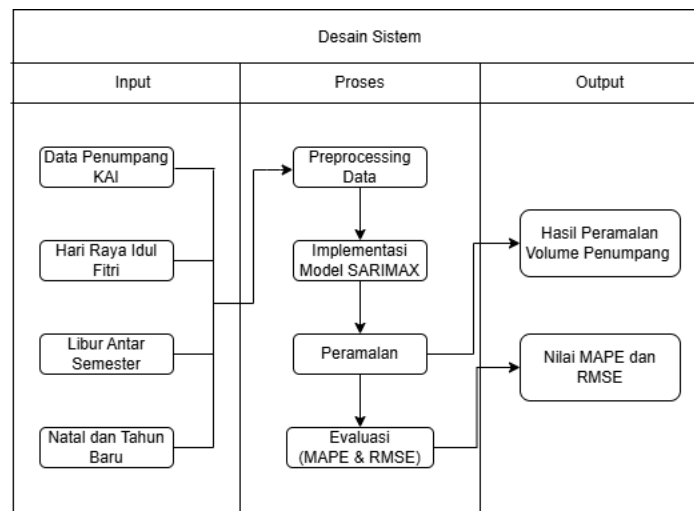
Absolute Percentage Error (MAPE) guna mengetahui tingkat akurasi model yang dihasilkan.

3.2 Pengumpulan Data

Informasi data yang dikumpulkan berkaitan dengan jumlah penumpang KAI di Pulau Jawa dari Januari 2020 sampai Desember 2025 dan daftar hari libur yaitu hari raya Idul Fitri, libur sekolah, Nataru. Data penumpang KAI diambil melalui *website* resmi BPS (Badan Pusat Statistik) sedangkan data hari libur diolah secara manual menjadi variabel eksogen biner, di mana nilai 1 diberikan untuk bulan yang memiliki hari libur/hari raya besar (seperti Idul Fitri, libur antar semester, Natal, dan Tahun Baru) yang berdampak signifikan pada lonjakan penumpang, dan nilai 0 untuk bulan yang tidak memiliki momen tersebut.

3.3 Desain Sistem

Desain sistem menjelaskan alur pengolahan data jumlah penumpang kereta api di Pulau Jawa dalam proses peramalan. Perancangan sistem ini disusun untuk memberikan gambaran yang jelas mengenai setiap tahapan pengolahan data sekaligus memudahkan implementasi metode SARIMAX dalam menyelesaikan permasalahan penelitian.



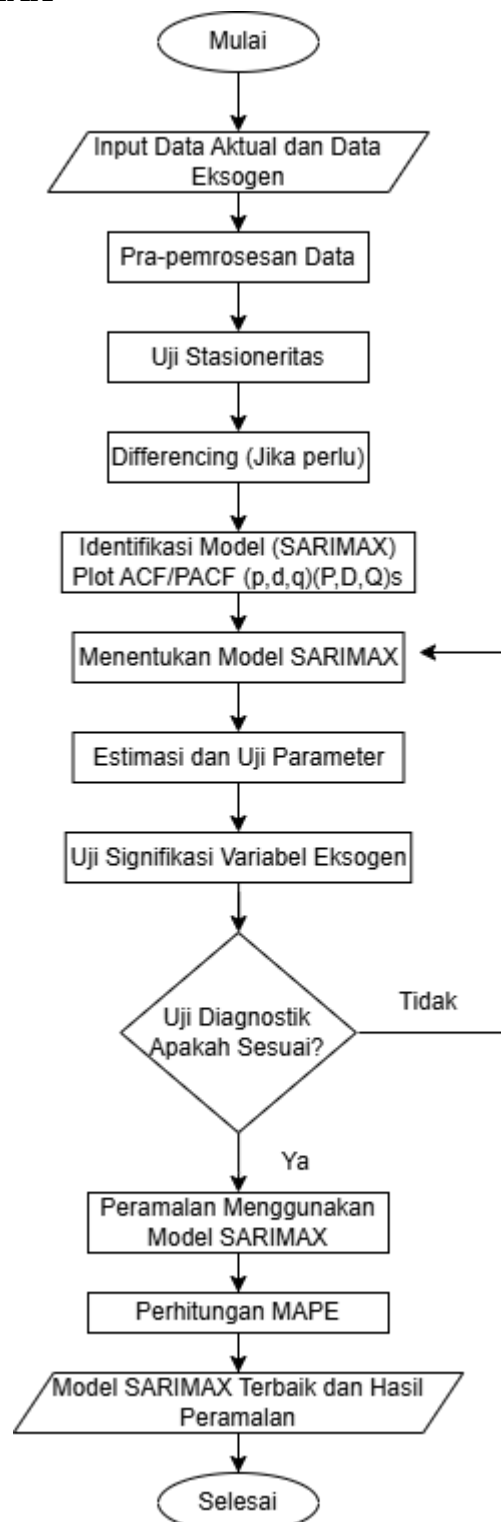
Gambar 3.2 Desain Sistem

Gambar 3.2 tersebut menggambarkan desain sistem yang digunakan dalam penelitian untuk melakukan peramalan volume penumpang kereta api di Pulau Jawa menggunakan metode SARIMAX. Tahap pertama adalah input, yang terdiri atas data volume penumpang kereta api sebagai variabel utama serta tiga variabel eksogen, yaitu Hari Raya Idul Fitri, libur antar semester, dan Natal serta Tahun Baru. Variabel eksogen tersebut dipilih karena memiliki pengaruh terhadap perubahan jumlah penumpang pada periode tertentu. Seluruh data kemudian diproses melalui tahap *preprocessing* untuk memastikan kualitas data, sekaligus membentuk variabel *dummy* yang akan digunakan sebagai masukan dalam model SARIMAX.

Setelah proses *preprocessing* selesai, data digunakan pada tahap implementasi model SARIMAX untuk membangun model peramalan yang mempertimbangkan pola historis dan pengaruh faktor eksternal secara bersamaan. Model yang telah terbentuk selanjutnya digunakan untuk menghasilkan prediksi volume penumpang pada periode mendatang. Tahap terakhir adalah evaluasi

menggunakan metrik MAPE dan RMSE untuk mengukur tingkat akurasi serta besarnya kesalahan hasil prediksi. Dari keseluruhan proses tersebut dihasilkan dua keluaran utama, yaitu hasil peramalan volume penumpang kereta api dan nilai MAPE serta RMSE sebagai indikator kinerja model.

3.4 Metode SARIMAX



Gambar 3.3 Flowchart Proses Metode SARIMAX

Gambar 3.3 tersebut menggambarkan alur penelitian peramalan menggunakan metode SARIMAX, yang dimulai dari tahap input data aktual dan data eksogen, dilanjutkan dengan pra-pemrosesan serta uji stasioneritas. Jika data belum stasioner, dilakukan proses *differencing* (non-musiman dan musiman) sebelum melangkah ke tahap identifikasi model menggunakan plot ACF/PACF untuk menentukan orde $(p,d,q)(P,D,Q)s$. Setelah model ditentukan, dilakukan estimasi parameter dan uji signifikansi variabel eksogen, yang kemudian dievaluasi melalui uji diagnostik residual. Hasil pengujian ini digunakan untuk menilai kesesuaian model serta mengidentifikasi kemungkinan adanya pola yang belum dapat dijelaskan oleh model. Selanjutnya, model terbaik dipilih berdasarkan hasil evaluasi secara keseluruhan, termasuk nilai AIC, signifikansi parameter, dan performa peramalan.

3.4.1 Input Data Aktual dan Data Eksogen

Langkah awal dalam penelitian ini adalah mengintegrasikan data deret waktu utama dengan variabel eksternal. Karena SARIMAX merupakan model regresi yang berbasis deret waktu, variabel hari libur harus ditransformasi menjadi variabel biner (*dummy*). Hari libur dianggap sebagai kejutan (*shocks*) yang memengaruhi fluktuasi penumpang di luar pola musiman internal data. Variabel eksogen diberi nilai 1 untuk periode yang memiliki hari libur/cuti bersama signifikan dan nilai 0 untuk periode normal. Adapun data aktual jumlah penumpang dan data eksogen tersebut disajikan dalam tabel-tabel berikut.

Tabel 3.1 Jumlah Penumpang KAI Tahun 2020-2025

Bulan	2020	2021	2022	2023	2024	2025
Januari	33.472	11.631	17.499	28.435	33.590	35.286
Februari	31.679	11.223	12.857	25.744	31.397	34.271
Maret	22.949	13.995	19.133	29.249	32.277	33.736
April	5.813	14.590	19.436	27.731	34.241	36.902
Mei	5.476	14.627	23.114	29.886	34.665	36.758
Juni	9.271	14.313	23.160	29.579	34.524	37.211
Juli	12.205	5.646	25.272	32.372	37.422	40.484
Agustus	12.679	6.479	23.920	30.742	35.292	36.694
September	11.295	9.510	25.167	31.067	34.792	36.376
Oktober	11.768	13.139	27.143	33.108	37.233	39.704
November	13.523	15.161	26.473	32.424	34.849	38.343
Desember	13.262	17.207	29.245	34.500	37.537	41.208

Tabel 3.2 Matriks Frekuensi Hari Libur (Tahun 2020-2025)

Tahun	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Ags	Sep	Okt	Nov	Des
2020	0	0	0	0	0	0	0	0	0	0	0	0
2021	0	0	0	0	0	0	0	0	0	0	0	0
2022	1	0	0	0	1	1	1	0	0	0	0	1
2023	1	0	0	1	1	1	1	0	0	0	0	1
2024	1	0	0	1	0	1	1	0	0	0	0	1
2025	1	0	1	1	0	1	1	0	0	0	0	1

3.4.2 Pra-Pemrosesan Data

Tahapan pra-pemrosesan data dilakukan untuk memastikan bahwa data yang telah terkumpul memiliki kualitas yang memadai, tersusun secara konsisten, dan siap digunakan dalam penyusunan model SARIMAX. Pada fase ini, berbagai langkah persiapan diterapkan secara terstruktur agar data dapat dianalisis secara optimal pada tahap pemodelan selanjutnya.

3.4.2.1 Data Cleaning

Tahap awal dilakukan dengan memverifikasi keberadaan nilai yang hilang (*missing values*) maupun potensi inkonsistensi dalam pencatatan data jumlah penumpang kereta api selama periode 2020-2025. Hasil pemeriksaan menunjukkan

bahwa seluruh rangkaian data tersedia secara utuh sebanyak 72 observasi bulanan, sehingga tidak diperlukan penerapan metode imputasi untuk melengkapi data.

3.4.2.2 Date Formatting

Data yang semula disajikan dalam bentuk matriks tahunan sebagaimana terlihat pada Tabel 3.1 kemudian ditransformasikan menjadi satu kolom deret waktu yang tersusun secara vertikal. Transformasi ini dilakukan agar model SARIMAX dapat mengidentifikasi dan memproses urutan waktu secara berkesinambungan mulai dari periode t_1 sampai dengan t_{72} .

Tabel 3.3 Data Transformasi Jumlah Penumpang KAI

<i>Date</i>	Penumpang
Jan-20	33472
Feb-20	31679
Mar-20	22949
Apr-20	5813
May-20	5476
...	...
Des-25	41208

3.4.2.3 Integrasi Variabel Eksogen

Pada tahap ini, data deret waktu jumlah penumpang (Y_t) digabungkan dengan variabel *dummy* hari libur (X_t) yang telah disusun sebelumnya. Setiap observasi penumpang dipasangkan dengan nilai indikator 1 atau 0 berdasarkan ada atau tidaknya perayaan hari raya/hari libur pada bulan terkait. Hasil penggabungan tersebut membentuk satu dataset yang selanjutnya digunakan sebagai dasar dalam proses estimasi parameter model.

Tabel 3.4 Integrasi Variabel Eksogen

<i>Date</i>	Penumpang	hari_libur
Jan-20	33472	0
Feb-20	31679	0
Mar-20	22949	0
Apr-20	5813	0
May-20	5476	0
...	...	...
Des-25	41208	1

3.4.3 Uji Stasioneritas

Dalam penelitian ini, pengujian kestasioneran data dilakukan menggunakan metode *Augmented Dickey-Fuller* (ADF) sebagai pengujian formal untuk mengetahui apakah data telah memenuhi asumsi stasioner sebelum proses pemodelan dilakukan. Uji ADF dilaksanakan dengan membandingkan nilai statistik uji terhadap nilai kritis pada taraf signifikansi 5% ($\alpha = 0,05$). Pengujian ini menggunakan hipotesis sebagai berikut.

H_0 : Data tidak stasioner

H_1 : Data stasioner

Keputusan dalam uji ADF ditentukan berdasarkan nilai *p-value*. Apabila nilai *p-value* lebih kecil dari 0,05, maka H_0 ditolak sehingga dapat disimpulkan bahwa data telah bersifat stasioner dan memenuhi syarat untuk digunakan pada tahap pemodelan selanjutnya.

3.4.4 Differencing

Sebelum membangun model SARIMAX, data deret waktu harus memenuhi asumsi stasioner agar hasil pemodelan lebih akurat. Salah satu teknik yang umum digunakan untuk mencapai kondisi tersebut adalah *differencing*, yaitu proses menghitung selisih antara nilai pada suatu periode dengan nilai pada periode

sebelumnya. Teknik ini bertujuan untuk mengurangi pengaruh tren maupun pola musiman yang menyebabkan data belum stasioner. Pada penelitian ini, proses *differencing* dilakukan melalui dua tahap, yaitu *non-seasonal differencing* untuk mengatasi tren dan *seasonal differencing* untuk menghilangkan pengaruh pola musiman yang berulang.

3.4.4.1 Non-Seasonal Differencing (d)

Non-seasonal differencing diterapkan untuk menghilangkan kecenderungan tren linear pada data, yaitu pola kenaikan atau penurunan yang berlangsung dalam jangka panjang. Tahapan ini dilakukan dengan menghitung perbedaan antara nilai observasi pada period ke-t dan nilai pada periode sebelumnya (t-1). Secara matematis, proses tersebut dapat dinyatakan melalui rumus berikut.

$$\Delta Y_t = Y_t - Y_{t-1} \quad (3.1)$$

Dimana:

- ΔY_t : Data hasil diferensiasi pada periode t.
- Y_t : Data aktual pada periode t.
- Y_{t-1} : Data aktual pada periode satu bulan sebelumnya.

3.4.4.2 Seasonal Differencing (D)

Mengingat data jumlah penumpang menunjukkan pola yang berulang setiap tahun, maka diterapkan diferensiasi musiman untuk menghilangkan unsur musiman tersebut. Karena data yang dianalisis berbentuk bulanan, nilai lag musiman yang digunakan adalah 12 periode. Secara matematis, prosedur ini dapat dinyatakan dengan rumus berikut:

$$\Delta_{12} Y_t = Y_t - Y_{t-12} \quad (3.2)$$

Dimana:

- ΔY_{t-12} : Data aktual pada bulan yang sama di tahun sebelumnya.

3.4.4.3 Penentuan Orde Diferensiasi

Banyaknya tahapan diferensiasi yang diterapkan akan menentukan nilai orde d untuk komponen nonmusiman dan D untuk komponen musiman dalam model SARIMAX. Apabila kestasioneran tercapai setelah satu kali diferensiasi biasa, maka ditetapkan $d = 1$, sedangkan jika kestasioneran diperoleh melalui satu kali diferensiasi musiman, maka $D = 1$. Prosedur ini dilakukan secara bertahap hingga grafik deret waktu tidak lagi memperlihatkan kecenderungan tren yang jelas dan hasil uji ADF menunjukkan nilai p -value di bawah 0,05. Data yang telah memenuhi kondisi tersebut selanjutnya digunakan untuk menentukan orde AR dan MA melalui analisis plot ACF dan PACF.

3.4.5 Identifikasi Model SARIMAX

Setelah data mencapai kondisi stasioner melalui proses *differencing*, langkah berikutnya adalah mengidentifikasi parameter model SARIMAX menggunakan plot ACF dan PACF. Plot ACF digunakan untuk menentukan orde *moving average* nonmusiman (q), sedangkan plot PACF digunakan untuk mengidentifikasi orde *autoregressive* nonmusiman (p). Selain itu, pola yang muncul pada *lag* musiman dimanfaatkan untuk menentukan parameter musiman, yaitu *seasonal autoregressive* (P) dan *seasonal moving average* (Q). Dengan demikian, hasil identifikasi dari kedua plot tersebut menjadi dasar dalam menentukan kombinasi parameter awal yang akan digunakan pada proses pemodelan SARIMAX.

3.4.6 Menentukan Model

Penentuan model yang akan digunakan tidak hanya mengandalkan model yang ditemukan berdasarkan plot ACF dan PACF saja. Untuk menghindari adanya kombinasi model yang terlewat, kombinasi parameter di sekitar orde yang sudah didapatkan sebelumnya akan digunakan dalam proses *grid search* untuk memperoleh model SARIMAX terbaik. Berdasarkan hasil dari *grid search* tersebut maka akan diambil 10 model terbaik dengan AIC terkecil.

3.4.7 Estimasi dan Uji Parameter

Langkah estimasi parameter diimplementasikan untuk menentukan besaran nilai koefisien pada model, yang mencakup komponen *Autoregressive* (AR), *Moving Average* (MA), *Seasonal AR*, *Seasonal MA*, serta variabel eksogen hari libur. Secara sistematis, proses penentuan angka-angka tersebut cukup kompleks karena melibatkan penentuan pemilihan nilai awal yang kemudian dimodifikasi secara berulang melalui proses iterative hingga mencapai titik optimal. Oleh sebab itu, pengerjaan ini dibantu oleh perangkat lunak computer dengan menerapkan metode *Maximum Likelihood Estimator* (MLE). Melalui penggunaan metode tersebut, diperoleh nilai parameter yang mampu merepresentasikan pola data penumpang secara akurat dalam struktur model yang telah dibentuk.

3.4.8 Uji Diagnostik

Tahap uji diagnostik diimplementasikan untuk menjamin validitas serta menentukan model optimal melalui serangkaian pengujian kelayakan. Prosedur ini difokuskan pada analisis residual guna memastikan bahwa seluruh informasi pada

data telah tertangkap secara sempurna oleh model. Penilaian terhadap residual tersebut dilakukan melalui verifikasi kondisi *white noise* yang secara spesifik diuji menggunakan metode *Ljung-Box* untuk mendeteksi adanya autokorelasi pada residual. Hipotesis dari uji *Ljung-Box* adalah sebagai berikut:

H_0 = residual memenuhi *white noise*

H_1 = residual tidak memenuhi *white noise*

3.4.9 Skenario Pengujian

Pada tahap ini dilakukan beberapa skenario pengujian untuk mengevaluasi kinerja model dalam menghasilkan peramalan. Pengujian dilakukan dengan membagi dataset menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih digunakan untuk membangun model sekaligus mempelajari pola yang terdapat pada data historis. Oleh karena itu, pemilihan data latih yang sesuai menjadi langkah penting karena akan memengaruhi kemampuan model dalam mengenali pola data dan menghasilkan prediksi yang lebih akurat pada tahap pengujian.

Setelah proses pelatihan selesai, model kemudian diuji menggunakan data uji (*testing data*) untuk mengetahui kemampuannya dalam memprediksi data yang belum pernah digunakan selama proses pelatihan. Pada penelitian ini diterapkan tiga skenario pembagian data, yaitu 90% data latih dan 10% data uji, 80% data latih dan 20% data uji, serta 70% data latih dan 30% data uji. Penerapan beberapa skenario tersebut bertujuan untuk membandingkan kinerja model dan menentukan komposisi data yang menghasilkan performa peramalan terbaik.

Selanjutnya, proses peramalan dilakukan menggunakan model SARIMAX dengan parameter terbaik yang telah diperoleh pada tahap pemodelan. Nilai prediksi volume penumpang kereta api dihitung berdasarkan persamaan SARIMAX yang ditunjukkan pada Persamaan (2.3). Hasil prediksi dari setiap skenario kemudian dibandingkan untuk mengevaluasi tingkat akurasi model dan menentukan konfigurasi yang memberikan hasil peramalan paling optimal.

Setiap skenario pembagian data selanjutnya dievaluasi menggunakan metrik MAPE dan RMSE dengan mengacu pada rumus yang terdapat pada Persamaan (2.5) dan Persamaan (2.6). Kedua metrik tersebut digunakan untuk mengukur tingkat kesalahan hasil peramalan dengan membandingkan nilai prediksi terhadap data aktual sehingga dapat diketahui kemampuan model dalam menghasilkan prediksi yang mendekati kondisi sebenarnya. Semakin kecil nilai MAPE dan RMSE yang diperoleh, semakin baik pula performa model dalam melakukan peramalan. Oleh karena itu, skenario yang menghasilkan nilai kesalahan paling rendah dipilih sebagai model yang paling optimal dalam penelitian ini.

Tabel 3.5 Skenario Pengujian

Skenario	Rasio Data Latih	Rasio Data Uji	Jumlah Data Latih	Jumlah Data Uji
Skenario 1	90%	10%	64	8
Skenario 2	80%	20%	57	15
Skenario 3	70%	30%	50	22

Pengujian model dilakukan menggunakan tiga skenario pembagian data, yaitu 90:10, 80:20, 70:30. Tujuan penggunaan beberapa skenario adalah untuk mengetahui pengaruh proporsi data latih dan data uji terhadap performa model SARIMAX dalam melakukan peramalan. Pada setiap skenario, model dilatih menggunakan data latih kemudian dievaluasi menggunakan data uji. Kinerja model

pada masing-masing skenario dibandingkan berdasarkan nilai MAPE dan RMSE, sehingga dapat ditentukan skenario pembagian data yang menghasilkan tingkat akurasi terbaik.

BAB IV

HASIL DAN PEMBAHASAN

Bab ini membahas implementasi metode SARIMAX untuk meramalkan volume penumpang kereta api di Pulau Jawa. Pada tahap ini dilakukan pengujian terhadap beberapa kombinasi model SARIMAX yang telah ditentukan berdasarkan hasil identifikasi parameter pada tahap sebelumnya. Setiap model dievaluasi untuk mengetahui kemampuannya dalam menghasilkan prediksi yang akurat. Proses evaluasi dilakukan dengan membandingkan nilai MAPE dan RMSE dari masing-masing model sehingga dapat diketahui tingkat kesalahan peramalannya. Model yang menghasilkan nilai error paling rendah dipilih sebagai model terbaik dan digunakan pada proses peramalan. Hasil dari seluruh pengujian model SARIMAX tersebut disajikan pada pembahasan berikut.

4.1 Uji Stasioneritas

Langkah pertama dalam penelitian ini adalah menguji kestasioneran data menggunakan uji *Augmented Dickey-Fuller* (ADF). Pengujian ini dilakukan untuk memastikan bahwa data telah memenuhi asumsi stasioneritas yang menjadi syarat dalam pemodelan deret waktu. Data dinyatakan stasioner apabila nilai *p-value* lebih kecil dari 0,05. Sebaliknya, jika nilai *p-value* lebih besar dari 0,05, maka data masih dianggap tidak stasioner sehingga perlu dilakukan proses transformasi, seperti *differencing*, sebelum digunakan pada tahap pemodelan. Hasil uji ADF terhadap data volume penumpang kereta api di Pulau Jawa ditampilkan pada Tabel 4.1.

Tabel 4.1 Uji Stasioner

Tes ADF (p-value)	Tanda	α	Keputusan	Kesimpulan
0,89	>	0,05	H ₀ diterima	Tidak Stasioner

Berdasarkan hasil yang ditunjukkan pada Tabel 4.1, data volume penumpang kereta api di Pulau Jawa belum memenuhi asumsi stasioneritas. Hal ini ditunjukkan oleh nilai *p-value* uji ADF sebesar 0,89, yang masih lebih besar dari batas signifikansi 0,05. Oleh karena itu, data perlu melalui proses *differencing* untuk mengurangi pengaruh tren sehingga karakteristik data menjadi lebih stabil. Tahap ini perlu dilakukan agar data memenuhi syarat untuk digunakan pada proses pemodelan SARIMAX.

4.1.1 Non-Seasonal Differencing (d)

Sebagai langkah awal, dilakukan *differencing* pertama pada data jumlah penumpang KAI di Pulau Jawa. Apabila hasil *differencing* pertama belum menghasilkan data yang stasioner, maka proses dilanjutkan dengan *differencing* kedua. Pada tahap *non-seasonal differencing*, nilai jumlah penumpang pada periode ke-*t* dikurangi dengan nilai pada periode sebelumnya. Proses perhitungan dilakukan menggunakan rumus yang mengacu pada Persamaan (3.1). Hasil *non-seasonal differencing* pertama ditampilkan pada Tabel 4.2.

Tabel 4.2 Uji Stasioner Differencing Pertama

Tes ADF (p-value)	Tanda	α	Keputusan	Kesimpulan
0,15	>	0,05	H ₀ diterima	Tidak Stasioner

Berdasarkan Tabel 4.2, data jumlah penumpang KAI di Pulau Jawa yang telah mengalami satu kali proses *differencing* masih belum stasioner. Hal ini ditunjukkan oleh nilai *p-value* uji ADF sebesar 0,15 yang masih berada di atas

ambang batas signifikansi 0,05. Oleh karena itu, dilakukan *differencing* tahap kedua untuk mencapai kondisi data yang lebih stabil. Hasil dari proses *differencing* kedua selanjutnya disajikan pada Tabel 4.3 sebagai dasar untuk mengevaluasi kembali tingkat stasioneritas data.

Tabel 4.3 Uji Stasioneritas *Differencing* Kedua

Tes ADF (p-value)	Tanda	α	Keputusan	Kesimpulan
0,0019	<	0,05	H ₀ ditolak	Stasioner

Berdasarkan hasil pada Tabel 4.3, data volume penumpang kereta api di Pulau Jawa telah memenuhi asumsi stasioner setelah dilakukan dua kali proses *non-seasonal differencing*. Hal ini ditunjukkan oleh nilai *p-value* uji ADF sebesar 0,0019, yang lebih kecil dari tingkat signifikansi 0,05. Hasil tersebut menunjukkan bahwa data sudah berada dalam kondisi stasioner sehingga proses transformasi tidak perlu dilanjutkan ke tahap *differencing* berikutnya dan data dapat digunakan pada tahap pemodelan SARIMAX.

4.1.2 Seasonal Differencing

Langkah selanjutnya yaitu melakukan *seasonal differencing* untuk menghilangkan unsur musiman yang ada pada data. *Seasonal differencing* dilakukan dengan mengurangi data jumlah penumpang pada periode ke-*t* dan data jumlah penumpang pada periode ke-*t* di tahun sebelumnya. Perhitungan dilakukan dengan menggunakan rumus yang merujuk pada persamaan (3.2). Hasil *seasonal differencing* pertama akan ditampilkan pada Tabel 4.4.

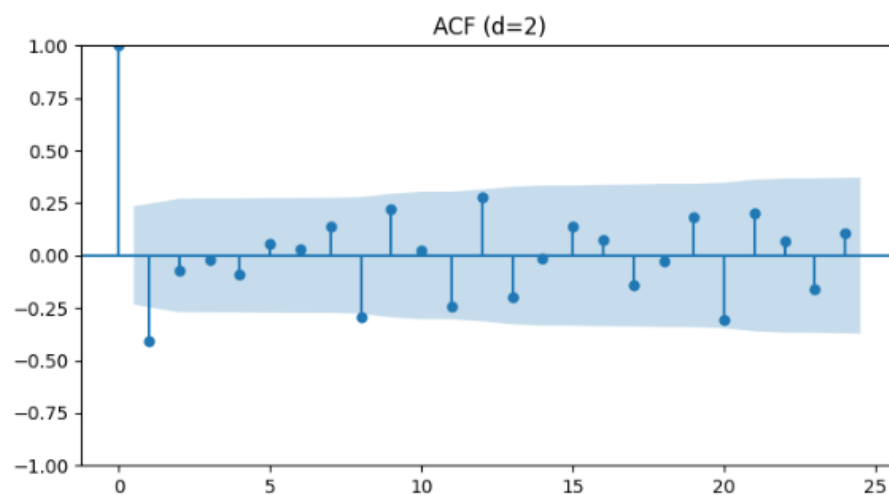
Tabel 4.4 Uji Stasioner *Seasonal Differencing*

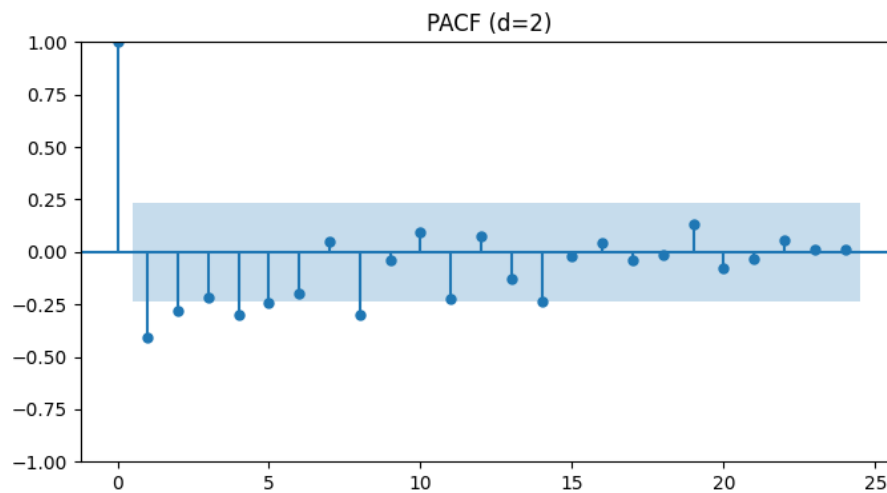
Tes ADF (p-value)	Tanda	α	Keputusan	Kesimpulan
0,0000	<	0,05	H ₀ ditolak	Stasioner

Berdasarkan hasil pada Tabel 4.4, data jumlah penumpang KAI yang telah melalui satu kali proses *differencing* sudah memenuhi asumsi stasioneritas. Hal ini terlihat dari nilai *p-value* uji ADF sebesar 0,0000 yang lebih kecil dari batas signifikansi 0,05, sehingga tidak perlu dilakukan proses *differencing* selanjutnya.

4.2 Identifikasi Model

Beberapa kandidat model yang akan diuji diperoleh dari hasil identifikasi parameter menggunakan plot ACF dan PACF pada data yang telah melalui dua kali *non-seasonal differencing* dan satu kali *seasonal differencing*. Sebagai dasar dalam proses identifikasi, hasil plot ACF dan PACF setelah *differencing* disajikan pada Gambar 4.1. Plot tersebut digunakan untuk menentukan kombinasi orde model yang akan diuji lebih lanjut pada tahap pembentukan model SARIMAX.





Gambar 4.1 Plot ACF dan PACF Setelah Differencing

Area berwarna biru pada plot ACF dan PACF yang terdapat pada Gambar 4.1 merupakan batas signifikansi (*significance level*) yang mewakili interval kepercayaan 95%. Batas tersebut berfungsi untuk menentukan apakah korelasi pada setiap lag bersifat signifikan secara statistik. Apabila batang korelasi melewati area biru tersebut, maka *lag* tersebut dianggap memiliki pengaruh signifikan terhadap model. Pada Gambar 4.1, plot ACF dan PACF digunakan sebagai dasar untuk menentukan parameter awal model SARIMAX. Berdasarkan plot ACF, parameter *moving average* nonmusiman (q) ditentukan dari *lag* yang melewati batas signifikansi. Hasil pengamatan menunjukkan bahwa hanya batang pada *lag* 1 yang melampaui batas tersebut sehingga diperoleh nilai $q = 1$. Sementara itu, parameter *seasonal moving average* (Q) ditentukan berdasarkan pola pada *lag* musiman, yaitu kelipatan 12. Pada plot ACF, batang di *lag* 12 berada tepat di sekitar batas signifikansi dan tidak diikuti lonjakan yang signifikan pada *lag* 24. Oleh karena itu, nilai parameter Q ditetapkan sebesar 0.

Sedangkan plot PACF digunakan untuk menentukan nilai p dan P . Pada plot PACF terlihat adanya lonjakan yang signifikan pada *lag* 1 sampai melewati batas signifikansi sehingga ditetapkan nilai $p = 1$. Sedangkan nilai P ditentukan dengan menggunakan cara yang sama dengan menentukan nilai Q . Sehingga dapat ditentukan nilai $P = 0$ karena semua batang yang terdapat pada *lag* kelipatan 12 berada di dalam batas signifikansi. Untuk nilai d dan D didapatkan dari proses *non-seasonal differencing* dan *seasonal differencing* yang sudah dilakukan sebelumnya yaitu, *non-seasonal differencing* dilakukan sebanyak dua kali sehingga $d = 2$ dan *seasonal differencing* dilakukan sebanyak satu kali sehingga $D = 1$. Model sementara yang didapatkan adalah SARIMAX (1,2,1)(0,1,0).

Selanjutnya, untuk memastikan seluruh kemungkinan kombinasi parameter telah dipertimbangkan, dilakukan pengujian beberapa kombinasi orde p , d , q serta P , D , Q menggunakan pendekatan grid search. Setiap model yang dihasilkan kemudian dievaluasi berdasarkan nilai Akaike Information Criterion (AIC), yang dihitung menggunakan rumus pada Persamaan (2.4). Nilai AIC digunakan untuk menilai kualitas model dengan mempertimbangkan keseimbangan antara kemampuan model dalam menjelaskan data dan jumlah parameter yang digunakan. Model dengan nilai AIC paling kecil dipilih karena dianggap memiliki performa yang lebih baik. Berdasarkan hasil identifikasi parameter melalui plot ACF, PACF, dan pendekatan grid search, diperoleh beberapa kandidat model SARIMAX yang selanjutnya akan dievaluasi pada tahap estimasi parameter. Kombinasi model yang digunakan dalam penelitian ini disajikan pada Tabel 4.5.

Tabel 4.5 Model Sementara yang Digunakan

No	Model SARIMAX	AIC
1	SARIMAX (1,2,1)(0,1,0)	1058.221

No	Model SARIMAX	AIC
2	SARIMAX (2,1,2)(0,0,1)	1009.9224
3	SARIMAX (2,1,2)(1,0,1)	1011.4541
4	SARIMAX (0,1,2)(0,0,1)	1024.2714
5	SARIMAX (0,1,2)(1,0,1)	1025.1632
6	SARIMAX (1,1,2)(0,0,1)	1026.2056
7	SARIMAX (1,1,2)(1,0,1)	1026.2896
8	SARIMAX (2,1,2)(1,0,0)	1033.3361
9	SARIMAX (0,1,1)(0,0,1)	1042.1297
10	SARIMAX (0,1,1)(1,0,1)	1042.1391
11	SARIMAX (1,1,1)(1,0,1)	1043.4351

4.3 Estimasi dan Uji Signifikansi Parameter

Dalam penelitian ini, estimasi parameter model SARIMAX dilakukan menggunakan bantuan perangkat lunak *python*. Parameter yang diestimasi meliputi komponen non-musiman (*autoregressive* dan *moving average*), komponen musiman (*seasonal autoregressive* dan *seasonal moving average*), serta koefisien variabel eksogen yang digunakan dalam model. Model dinyatakan signifikan apabila seluruh parameter memiliki nilai *p-value* kurang dari 0,05.

Tabel 4.6 Hasil Uji Signifikansi Parameter Pada Setiap Model

No	Model	Parameter	Estimasi Parameter	<i>p-value</i>	Ket	Keputusan
1	(1,2,1)(0,1,0)	AR	0.0783	0.704	>	Tidak Signifikan
		MA	-0.9704	0.000	<	
		eks	2744.4286	0.224	>	
2	(2,1,2)(0,0,1)	AR1	1.0075	0.000	<	Tidak Signifikan
		AR2	-0.6947	0.000	<	
		MA1	-1.2725	0.000	<	
		MA2	0.9746	0.000	<	
		S MA	0.4667	0.038	<	
		eksogen	2304.7060	0.123	>	
3	(2,1,2)(1,0,1)	AR1	0.9860	0.000	<	Tidak Signifikan
		AR2	-0.6862	0.000	<	
		MA1	-1.2576	0.000	<	
		MA2	0.9780	0.000	<	
		S AR	-0.1069	0.633	>	
		S MA	0.5459	0.018	<	
3	(0,1,2)(0,0,1)	eksogen	2304.7060	0.110	>	Tidak Signifikan
		MA1	-0.1047	0.632	>	
		MA2	-0.0326	0.809	>	
		S MA	0.2787	0.001	<	
5	(0,1,2)(1,0,1)	eksogen	2304.7059	0.016	<	Tidak Signifikan
		MA1	-0.1104	0.639	>	
		MA2	-0.0344	0.823	>	

No	Model	Parameter	Estimasi Parameter	<i>p-value</i>	Ket	Keputusan
		S AR	-0.2328	0.378	>	
		S MA	0.5070	0.068	>	
		eksogen	2304.7058	0.029	<	
6	(1,1,2)(0,0,1)	AR	0.1721	0.908	>	Tidak Signifikan
		MA1	-0.2817	0.865	>	
		MA2	-0.0418	0.884	>	
		S MA	0.2836	0.001	<	
		eksogen	2304.7058	0.016	<	
7	(1,1,2)(1,0,1)	AR	-0.5429	0.335	>	Tidak Signifikan
		MA1	0.4612	0.462	>	
		MA2	-0.0005	0.998	>	
		S AR	-0.2427	0.351	>	
		S MA	0.5022	0.054	>	
		eksogen	2304.7057	0.045	<	
8	(2,1,2)(1,0,0)	AR1	0.6753	0.000	<	Signifikan
		AR2	-0.8380	0.000	<	
		MA1	-0.5943	0.005	<	
		MA2	0.9730	0.000	<	
		S AR	0.3046	0.011	<	
		eksogen	2304.7060	0.014	<	
9	(0,1,1)(0,0,1)	MA	-0.0816	0.112	>	Tidak Signifikan
		S MA	0.1286	0.679	>	
		eksogen	2304.7059	0.014	<	
10	(0,1,1)(1,0,1)	MA	-0.0911	0.690	>	Tidak Signifikan
		S AR	-0.2904	0.234	>	
		S MA	0.4691	0.097	>	
		eksogen	2304.7058	0.021	<	
11	(1,1,1)(1,0,1)	AR	-0.5466	0.401	>	Tidak Signifikan
		MA	0.4909	0.501	>	
		S AR	-0.3146	0.169	>	
		S MA	0.4926	0.045	<	
		eksogen	2304.7054	0.033	<	

Tabel 4.6 menyajikan hasil uji signifikansi parameter untuk seluruh model yang diuji. Parameter dengan nilai *p-value* kurang dari 0,05 dinyatakan signifikan, yang berarti memiliki kontribusi terhadap model. Sebaliknya, parameter dengan nilai *p-value* lebih besar dari 0,05 dianggap tidak signifikan karena pengaruhnya terhadap model tidak cukup kuat secara statistik.

Berdasarkan hasil pengujian terhadap 11 model, hanya ada satu model yang memenuhi kriteria signifikansi pada seluruh parameternya, yaitu SARIMAX (2,1,2)(1,0,0). Model tersebut memiliki nilai *p-value* di bawah 0,05 untuk setiap

parameter yang diestimasi, sehingga seluruh komponen dinilai berkontribusi dalam menjelaskan pola data. Sementara itu, model lainnya masih memiliki parameter yang tidak signifikan sehingga tidak dipilih untuk tahap selanjutnya.

4.4 Uji Kesesuaian Model

Model SARIMAX (2,1,2)(1,0,0) merupakan model yang seluruh parameternya signifikan dari 11 model yang telah mengalami uji signifikansi. Model tersebut akan digunakan untuk tahapan selanjutnya yaitu uji kesesuaian model. Tahap ini dilakukan untuk mengevaluasi karakteristik residual yang dihasilkan sebelum digunakan pada tahap peramalan.

Tabel 4.7 Hasil Uji *White Noise*

Model	Lag	<i>p-value</i>	Keterangan	Keputusan
(2,1,2)(1,0,0)	12	0,113359	> 0,05	Signifikan
	24	0,600111	> 0,05	Signifikan
	36	0,915127	> 0,05	Signifikan

Berdasarkan hasil uji Ljung-Box yang disajikan pada Tabel 4.7, diperoleh nilai *p-value* sebesar 0,113359 pada *lag* 12, kemudian 0,600111 pada *lag* 24, dan 0,915127 pada *lag* 36. Seluruh nilai tersebut berada di atas tingkat signifikansi 0,05 sehingga hipotesis 0 tidak dapat ditolak. Hasil pengujian ini menunjukkan bahwa residual pada model SARIMAX (2,1,2)(1,0,0) tidak memiliki autokorelasi yang signifikan. Dengan demikian, residual dapat dianggap memenuhi karakteristik *white noise*, sehingga asumsi residual pada model telah terpenuhi. Berdasarkan hasil tersebut, model SARIMAX (2,1,2)(1,0,0) dinilai layak untuk digunakan pada tahap peramalan.

4.5 Skenario Pengujian

Model SARIMAX (2,1,2)(1,0,0) selanjutnya akan digunakan dalam serangkaian skenario pengujian. Pengujian dilakukan dengan beberapa variasi pembagian data, yaitu rasio 90:10, 80:20, dan 70:30 antara data latih (*training data*) dan data uji (*testing data*). Variasi tersebut diterapkan untuk melihat performa model pada kondisi pembagian data yang berbeda.

Pengujian pertama menggunakan perbandingan 90% data latih dan 10% data uji. Hasil MAPE dan RMSE pengujian pertama akan ditampilkan pada tabel berikut.

Tabel 4.8 Pengujian MAPE Pertama

Periode	Data Aktual (Y_t)	Prediksi ($\hat{Y}_t$)	$Y_t - \hat{Y}_t$	$\frac{ Y_t - \hat{Y}_t }{Y_t}$
Mei 2025	36758	36940	-182	0,004
Juni 2025	37211	36829	382	0,010
Juli 2025	40484	37590	2894	0,071
Agustus 2025	36694	35499	1195	0,032
September 2025	36376	35408	968	0,026
Oktober 2025	39704	36010	3694	0,093
November 2025	38343	35340	3003	0,078
Desember 2025	41208	37633	3575	0,086

$$\begin{aligned}
 \text{MAPE} &= \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\% \\
 &= \frac{0.004 + 0.010 + 0.071 + 0.032 + 0.026 + 0.093 + 0.078 + 0.086}{8} \times 100\% \\
 &= 0.0505 \times 100\% \\
 &= 5.05\%
 \end{aligned}$$

Selanjutnya, akurasi model dievaluasi menggunakan nilai MAPE dengan membandingkan hasil prediksi terhadap data aktual pada periode pengujian, yaitu Mei 2025 hingga Desember 2025, sebagaimana ditunjukkan pada Tabel 4.8. Perhitungan MAPE dilakukan dengan menentukan selisih absolut antara nilai prediksi dan data aktual pada setiap periode, kemudian membaginya dengan nilai

aktual untuk memperoleh nilai *Absolute Percentage Error* (APE). Hasil perhitungan menunjukkan bahwa nilai APE berbeda pada setiap periode karena dipengaruhi oleh besarnya selisih antara hasil prediksi dan data aktual. Nilai APE terendah tercatat pada Mei 2025 sebesar 0,004, sedangkan nilai tertinggi terjadi pada Oktober 2025 sebesar 0,093. Temuan ini menunjukkan bahwa tingkat kesalahan prediksi model berubah pada setiap periode pengamatan, sehingga akurasi model juga dapat mengalami variasi dari waktu ke waktu.

Selanjutnya, seluruh nilai APE dijumlahkan kemudian dibagi dengan jumlah data pengujian, yaitu delapan periode, sehingga diperoleh nilai *Mean Absolute Percentage Error* (MAPE). Berdasarkan hasil perhitungan, diperoleh nilai MAPE sebesar 5,05%. Nilai tersebut menunjukkan bahwa rata-rata kesalahan prediksi model terhadap data aktual hanya sebesar 5,05%. Dengan kata lain, hasil prediksi yang dihasilkan model memiliki tingkat penyimpangan yang relatif kecil dibandingkan nilai aktual pada setiap periode pengujian. Hal ini menunjukkan bahwa model mampu mengikuti pola perubahan jumlah penumpang dengan cukup baik.

Berdasarkan klasifikasi tingkat akurasi MAPE, nilai sebesar 5,05% termasuk dalam kategori sangat akurat karena berada di bawah batas 10%. Hasil tersebut menunjukkan bahwa model SARIMAX (2,1,2)(1,0,0) mampu menghasilkan peramalan jumlah penumpang kereta api dengan tingkat ketepatan yang baik. Walaupun pada beberapa periode masih terdapat selisih antara nilai prediksi dan data aktual, perbedaannya relatif kecil sehingga tidak memengaruhi kinerja model secara keseluruhan. Temuan ini juga menunjukkan bahwa model

mampu mengikuti pola data dengan baik, baik dari sisi tren, pola musiman, maupun pengaruh variabel eksogen yang digunakan dalam penelitian.

Tabel 4.9 Pengujian RMSE Pertama

Periode	Data Aktual (Y_t)	Prediksi ($\hat{Y}_t$)	$Y_t - \hat{Y}_t$	$(Y_t - \hat{Y}_t)^2$
Mei 2025	36758	36940	-182	33124
Juni 2025	37211	36829	382	145924
Juli 2025	40484	37590	2894	8375236
Agustus 2025	36694	35499	1195	1428025
September 2025	36376	35408	968	937024
Oktober 2025	39704	36010	3694	13645636
November 2025	38343	35340	3003	9018009
Desember 2025	41208	37633	3575	12780625

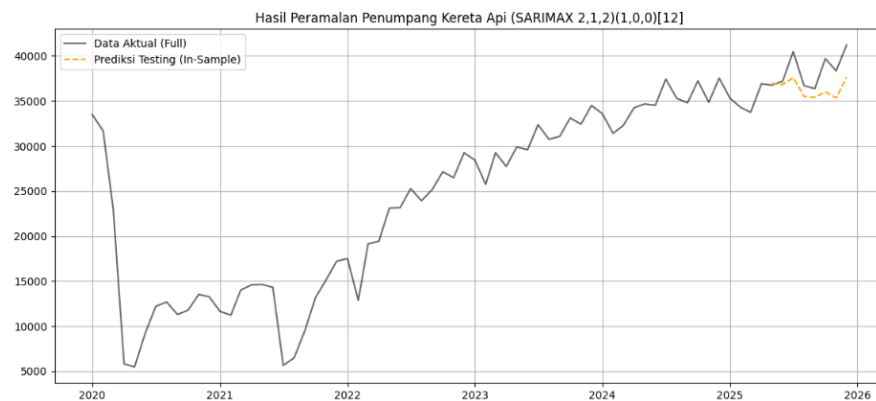
$$\begin{aligned}
 \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \\
 &= \sqrt{\frac{33124 + 145924 + 8375236 + 1428025 + 937024 + 13645636 + 9018009 + 12780625}{8}} \\
 &= \sqrt{\frac{46363603}{8}} \\
 &= 2407.38
 \end{aligned}$$

Selain menggunakan MAPE, kinerja model juga dievaluasi melalui nilai *Root Mean Square Error* (RMSE) yang disajikan pada Tabel 4.9. Nilai RMSE digunakan untuk mengetahui besarnya rata-rata kesalahan prediksi model terhadap data aktual. Proses perhitungannya dimulai dengan mencari selisih antara nilai prediksi dan data aktual pada setiap periode pengujian, kemudian setiap selisih dikuadratkan agar seluruh nilai kesalahan menjadi positif. Hasil perhitungan menunjukkan bahwa nilai kuadrat kesalahan terbesar terjadi pada Desember 2025, yaitu sebesar 12.780.625, sedangkan nilai terkecil diperoleh pada Mei 2025 sebesar 33.124. Perbedaan tersebut menunjukkan bahwa besarnya kesalahan prediksi

bervariasi pada setiap periode pengujian. Dengan demikian, nilai RMSE dapat memberikan gambaran mengenai tingkat kesalahan model secara keseluruhan.

Selanjutnya, seluruh nilai kuadrat kesalahan dijumlahkan kemudian dibagi dengan jumlah data pengujian, yaitu delapan periode. Hasil pembagian tersebut menghasilkan *Mean Squared Error* (MSE) yang selanjutnya diakarkan untuk memperoleh nilai RMSE. Berdasarkan perhitungan, diperoleh nilai RMSE sebesar 2.407,38. Nilai tersebut menunjukkan bahwa rata-rata penyimpangan hasil prediksi terhadap data aktual berada pada kisaran 2.407 penumpang untuk setiap periode pengujian. Semakin kecil nilai RMSE, maka semakin kecil pula kesalahan prediksi yang dihasilkan oleh model.

Nilai RMSE sebesar 2.407,38 menunjukkan bahwa model SARIMAX (2,1,2)(1,0,0) memiliki tingkat kesalahan prediksi yang relatif rendah jika dibandingkan dengan jumlah penumpang yang mencapai puluhan ribu setiap bulan. Hasil ini menunjukkan bahwa nilai prediksi yang dihasilkan model cenderung mendekati data aktual. Temuan tersebut juga sejalan dengan nilai MAPE sebesar 5,05% yang termasuk dalam kategori sangat akurat, sehingga semakin memperkuat bahwa model memiliki kinerja peramalan yang baik berdasarkan dua metrik evaluasi yang berbeda. Secara keseluruhan, model tidak hanya mampu menjaga tingkat kesalahan prediksi tetap rendah, tetapi juga dapat mengikuti pola perubahan jumlah penumpang selama periode pengujian dengan baik.



Gambar 4.2 Grafik Hasil Peramalan Rasio 90:10

Gambar 4.2 menunjukkan perbandingan antara data aktual jumlah penumpang kereta api dengan hasil prediksi model SARIMAX $(2,1,2)(1,0,0)$ pada data pengujian dengan rasio pembagian data 90:10. Berdasarkan grafik tersebut, terlihat bahwa pola prediksi yang dihasilkan model mampu mengikuti pergerakan data aktual dengan cukup baik. Meskipun terdapat sedikit perbedaan pada beberapa periode, secara umum model dapat menangkap arah tren dan perubahan jumlah penumpang yang terjadi pada data pengujian.

Pada beberapa periode terlihat bahwa garis prediksi berada sangat dekat dengan garis data aktual, sedangkan pada periode lainnya masih terdapat jarak di antara kedua garis tersebut. Perbedaan tersebut tampak lebih jelas ketika data aktual mengalami perubahan yang cukup besar dalam waktu yang singkat. Garis prediksi cenderung membentuk pola yang lebih halus dibandingkan garis data aktual sehingga beberapa kenaikan dan penurunan belum sepenuhnya diikuti oleh model. Meskipun demikian, pola naik dan turun pada hasil prediksi tetap memiliki arah yang serupa dengan pola yang ditunjukkan oleh data aktual. Secara keseluruhan, grafik menunjukkan bahwa hasil peramalan memiliki pola pergerakan yang sejalan dengan perubahan jumlah penumpang selama periode pengujian.

Pengujian selanjutnya akan menggunakan perbandingan 80% data *training* dan 20% data *testing*. Hasil MAPE dan RMSE pengujian kedua akan ditampilkan pada tabel berikut.

Tabel 4.10 Pengujian MAPE Kedua

Periode	Data Aktual (Y_t)	Prediksi ($\hat{Y}_t$)	$Y_t - \hat{Y}_t$	$\left \frac{Y_t - \hat{Y}_t}{Y_t} \right $
Oktober 2024	37233	35586	1647	0,044
November 2024	34849	35148	-299	0,008
Desember 2024	37537	36837	700	0,018
Januari 2025	35286	36549	-1263	0,035
Februari 2025	34271	34711	-440	0,012
Maret 2025	33736	35139	-1403	0,041
April 2025	36902	37212	-310	0,008
Mei 2025	36758	37039	-281	0,007
Juni 2025	37211	36882	329	0,008
Juli 2025	40484	37730	2754	0,068
Agustus 2025	36694	35795	899	0,024
September 2025	36376	35689	687	0,018
Oktober 2025	39704	35726	3978	0,100
November 2025	38343	35501	2842	0,074
Desember 2025	41208	37579	3629	0,088

$$\begin{aligned}
 \text{MAPE} &= \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\% \\
 &= \frac{0,044 + 0,008 + 0,018 + 0,035 + 0,012 + 0,041 + 0,008 \dots + 0,088}{15} \times 100\% \\
 &= 0.0375 \times 100\% \\
 &= 3.75 \%
 \end{aligned}$$

Berdasarkan Tabel 4.10, nilai *Mean Absolute Percentage Error* (MAPE) dihitung dengan membandingkan hasil prediksi model terhadap data aktual pada periode pengujian, yaitu Oktober 2024 hingga Desember 2025. Perhitungan dilakukan dengan mencari nilai kesalahan absolut antara data aktual dan hasil prediksi, kemudian membaginya dengan nilai aktual pada setiap periode sehingga diperoleh nilai *Absolute Percentage Error* (APE). Nilai APE yang dihasilkan menunjukkan tingkat kesalahan prediksi pada masing-masing bulan selama periode

pengujian. Nilai APE terkecil diperoleh pada November 2024 dan April 2025 sebesar 0,008, sedangkan nilai terbesar terjadi pada Oktober 2025 sebesar 0,100. Perbedaan tersebut menunjukkan bahwa tingkat kesalahan prediksi model bervariasi pada setiap periode sesuai dengan karakteristik data yang diamati.

Setelah nilai *Absolute Percentage Error* (APE) pada setiap periode diperoleh, seluruh nilai tersebut dijumlahkan lalu dibagi dengan jumlah data pengujian, yaitu sebanyak 15 periode. Dari proses tersebut, diperoleh nilai MAPE sebesar 3,75%. Nilai ini menunjukkan bahwa rata-rata kesalahan prediksi model terhadap data aktual hanya mencapai 3,75% dari nilai sebenarnya. Semakin rendah nilai MAPE yang dihasilkan, semakin baik pula tingkat akurasi model dalam melakukan peramalan. Dengan hasil tersebut, dapat diketahui bahwa prediksi yang dihasilkan model memiliki kedekatan yang tinggi dengan data aktual selama periode pengujian.

Berdasarkan klasifikasi tingkat akurasi MAPE, nilai sebesar 3,75% termasuk dalam kategori sangat akurat karena berada di bawah batas 10%. Hasil ini menunjukkan bahwa model SARIMAX (2,1,2)(1,0,0) mampu mempelajari pola historis data serta pengaruh variabel eksogen dengan baik sehingga menghasilkan prediksi yang mendekati nilai aktual. Jika dibandingkan dengan skenario pembagian data lainnya, rasio 80:20 menghasilkan nilai MAPE paling rendah, sehingga menunjukkan performa peramalan yang lebih baik. Hal tersebut mengindikasikan bahwa rasio 80:20 memberikan keseimbangan yang optimal antara jumlah data pelatihan dan data pengujian sehingga model dapat mempelajari karakteristik data secara lebih efektif.

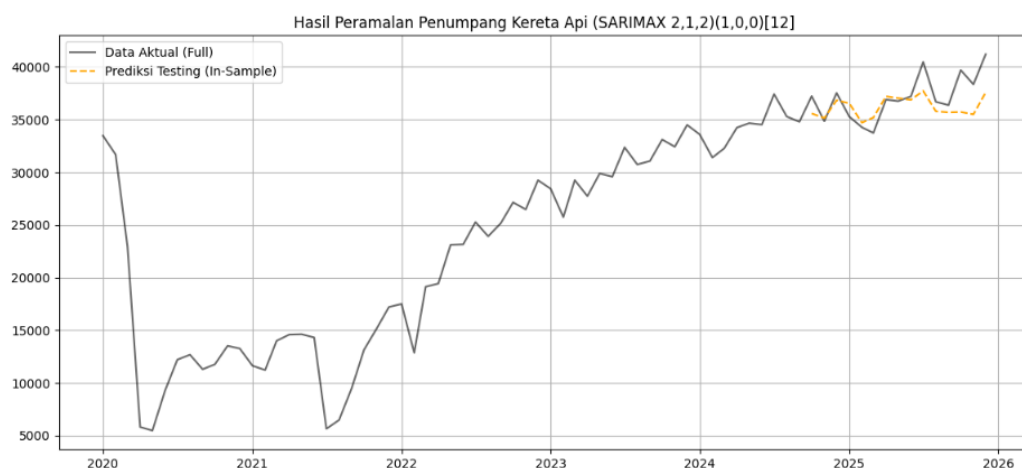
Tabel 4.11 Pengujian RMSE Kedua

Periode	Data Aktual (Y_t)	Prediksi ($\hat{Y}_t$)	$Y_t - \hat{Y}_t$	$(Y_t - \hat{Y}_t)^2$
Oktober 2024	37233	35586	1647	2712609
November 2024	34849	35148	-299	89401
Desember 2024	37537	36837	700	490000
Januari 2025	35286	36549	-1263	1595169
Februari 2025	34271	34711	-440	193600
Maret 2025	33736	35139	-1403	1968409
April 2025	36902	37212	-310	96100
Mei 2025	36758	37039	-281	78961
Juni 2025	37211	36882	329	108241
Juli 2025	40484	37730	2754	7584516
Agustus 2025	36694	35795	899	808201
September 2025	36376	35689	687	471969
oktober 2025	39704	35726	3978	15824484
November 2025	38343	35501	2842	8076964
Desember 2025	41208	37579	3629	13169641

$$\begin{aligned}
 \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \\
 &= \sqrt{\frac{1102500 + 8836 + 1052676 + 27556 + 3279721 + 3279721 + \dots + 66732561}{15}} \\
 &= \sqrt{\frac{53268265}{15}} \\
 &= 1887.12
 \end{aligned}$$

Berdasarkan Tabel 4.11, nilai kuadrat kesalahan (*Squared Error*) pada setiap periode menunjukkan variasi yang cukup beragam. Nilai kesalahan terkecil terjadi pada Mei 2025 sebesar 78.961, sedangkan kesalahan terbesar terjadi pada Desember 2025 sebesar 13.169.641. Perbedaan tersebut menunjukkan bahwa tingkat kedekatan hasil prediksi dengan data aktual berbeda pada setiap periode pengujian. Meskipun demikian, sebagian besar nilai kesalahan masih berada pada kisaran yang relatif rendah dibandingkan dengan jumlah penumpang yang mencapai puluhan ribu orang setiap bulannya. Hal ini menunjukkan bahwa model mampu mengikuti pola perubahan jumlah penumpang dengan cukup baik.

Selanjutnya, hasil perhitungan menunjukkan bahwa nilai RMSE yang diperoleh sebesar 1.887,12. Nilai tersebut mengindikasikan bahwa rata-rata selisih antara hasil prediksi dan data aktual berada pada kisaran 1.887 penumpang pada setiap periode pengujian. Jika dibandingkan dengan jumlah penumpang aktual yang berada pada rentang 33 ribu hingga 41 ribu penumpang, besarnya kesalahan tersebut masih tergolong rendah. Hal ini menunjukkan bahwa prediksi yang dihasilkan model cukup dekat dengan kondisi aktual sehingga tingkat kesalahan yang terjadi masih dapat diterima. Berdasarkan hasil tersebut, dapat disimpulkan bahwa model memiliki kemampuan yang baik dalam memprediksi volume penumpang kereta api.



Gambar 4.3 Grafik Hasil Peramalan Rasio 80:20

Gambar 4.3 menunjukkan grafik hasil peramalan jumlah penumpang kereta api pada skenario 80:20. Pada rasio pembagian data 80:20 menunjukkan bahwa garis prediksi mampu mengikuti pola pergerakan data aktual dengan cukup baik selama periode pengujian. Hal ini terlihat dari kedekatan antara garis prediksi dan garis data aktual pada sebagian besar periode pengamatan. Meskipun terdapat beberapa selisih ketika jumlah penumpang mengalami kenaikan atau penurunan

yang cukup tajam, model tetap mampu mengikuti arah perubahan data secara konsisten. Kondisi tersebut menunjukkan bahwa model mampu menangkap pola tren, pola musiman, serta pengaruh variabel eksogen secara lebih optimal.

Berdasarkan Gambar 4.3, garis prediksi cenderung memiliki perubahan yang lebih halus dibandingkan garis data aktual. Hal ini menyebabkan kenaikan dan penurunan pada data aktual belum sepenuhnya diikuti oleh hasil prediksi model. selisih tersebut terlihat pada saat jumlah penumpang mengalami peningkatan yang cukup tinggi maupun ketika terjadi penurunan pada periode tertentu. Namun, garis prediksi tetap berada di sekitar garis data aktual sehingga pola umum pergerakan data masih dapat direpresentasikan oleh model. gambaran tersebut menunjukkan bahwa hasil peramalan memiliki kecenderungan yang serupa dengan data aktual selama periode pengujian.

Pengujian selanjutnya akan menggunakan perbandingan 70% data *training* dan 30% data *testing*. Hasil MAPE pengujian ketiga akan ditampilkan pada tabel berikut.

Tabel 4.12 Pengujian MAPE Ketiga

Periode	Data Aktual (Y_t)	Prediksi ($\hat{Y}_t$)	$Y_t - \hat{Y}_t$	$\frac{ Y_t - \hat{Y}_t }{Y_t}$
Maret 2024	32277	31913	364	0,011
April 2024	34241	33307	934	0,027
Mei 2024	34665	34101	564	0,016
Juni 2024	34524	34164	360	0,010
Juli 2024	37422	34654	2768	0,073
Agustus 2024	35292	32403	2889	0,081
September 2024	34792	32458	2334	0,067
Oktober 2024	37233	33092	4141	0,111
November 2024	34849	33063	1786	0,051
Desember 2024	37537	35153	2384	0,063
Januari 2025	35286	34806	480	0,013
Februari 2025	34271	32586	1685	0,049
Maret 2025	33736	32780	956	0,028
April 2025	36902	34858	2044	0,055
Mei 2025	36758	35045	1713	0,046

Periode	Data Aktual (Y_t)	Prediksi ($\hat{Y}_t$)	$Y_t - \hat{Y}_t$	$\left \frac{Y_t - \hat{Y}_t}{Y_t} \right $
Juni 2025	37211	34983	2228	0,059
Juli 2025	40484	35039	5445	0,134
Agustus 2025	36694	32886	3808	0,103
September 2025	36376	32962	3414	0,093
Oktober 2025	39704	33131	6573	0,165
November 2025	38343	33087	5256	0,137
Desember 2025	41208	35185	6023	0,146

$$\begin{aligned}
 \text{MAPE} &= \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\% \\
 &= \frac{0,011 + 0,027 + 0,016 + 0,010 + 0,073 + 0,081 + 0,067 \dots + 0,146}{22} \times 100\% \\
 &= 0,0704 \times 100\% \\
 &= 7,04 \%
 \end{aligned}$$

Berdasarkan Tabel 4.12, nilai *Absolute Percentage Error* (APE) pada setiap periode pengujian menunjukkan adanya variasi tingkat kesalahan prediksi. Nilai APE terkecil diperoleh pada Juni 2024 sebesar 0,010, sedangkan nilai terbesar terjadi pada Oktober 2025 sebesar 0,165. Pada awal periode pengujian, nilai kesalahan cenderung rendah karena hasil prediksi masih mendekati data aktual. Namun, memasuki pertengahan hingga akhir periode pengujian terlihat bahwa selisih antara data aktual dan hasil prediksi semakin meningkat sehingga nilai APE menjadi lebih besar. Kondisi tersebut menunjukkan bahwa kemampuan model dalam mengikuti perubahan data mengalami penurunan pada beberapa periode terakhir.

Hasil evaluasi menunjukkan bahwa model menghasilkan nilai MAPE sebesar 7,04%. Nilai tersebut masih berada pada kategori sangat akurat karena kurang dari 10%, sehingga model masih mampu memberikan hasil prediksi yang baik. Meskipun demikian, nilai MAPE ini lebih besar dibandingkan hasil pengujian

pada rasio pembagian data 80:20 maupun 90:10. Perbedaan tersebut menunjukkan bahwa tingkat kesalahan prediksi model pada skenario ini relatif lebih tinggi. Hal ini mengindikasikan bahwa pembagian data 70:30 belum memberikan performa yang sebaik dua skenario pembagian data lainnya.

Tabel 4.13 Pengujian RMSE Ketiga

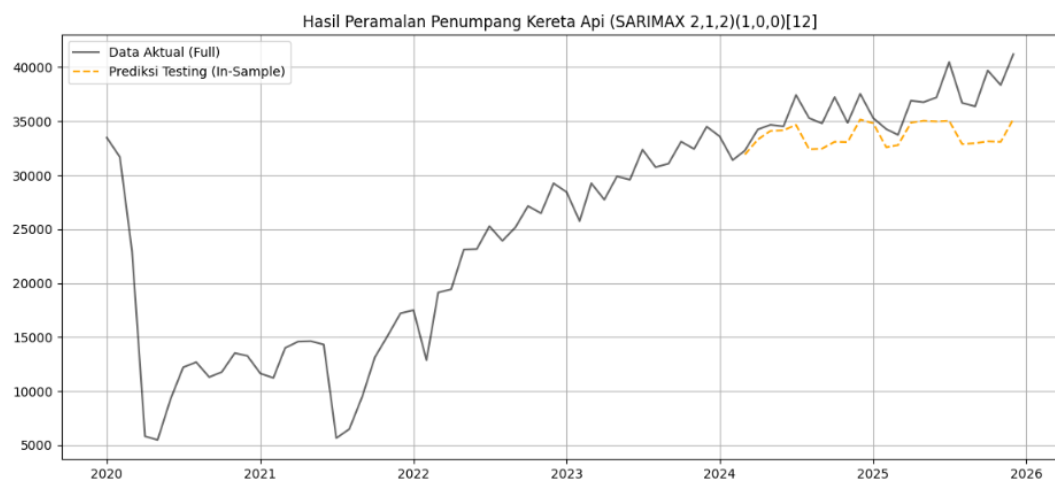
Periode	Data Aktual (Y_t)	Prediksi ($\hat{Y}_t$)	$Y_t - \hat{Y}_t$	$(Y_t - \hat{Y}_t)^2$
Maret 2024	32277	31913	364	132496
April 2024	34241	33307	934	872356
Mei 2024	34665	34101	564	318096
Juni 2024	34524	34164	360	129600
Juli 2024	37422	34654	2768	7661824
Agustus 2024	35292	32403	2889	8346321
September 2024	34792	32458	2334	5447556
Oktober 2024	37233	33092	4141	17147881
November 2024	34849	33063	1786	3189796
Desember 2024	37537	35153	2384	5683456
Januari 2025	35286	34806	480	230400
Februari 2025	34271	32586	1685	2839225
Maret 2025	33736	32780	956	913936
April 2025	36902	34858	2044	4177936
Mei 2025	36758	35045	1713	2934369
Juni 2025	37211	34983	2228	4963984
Juli 2025	40484	35039	5445	29648025
Agustus 2025	36694	32886	3808	14500864
September 2025	36376	32962	3414	11655396
Oktober 2025	39704	33131	6573	43204329
November 2025	38343	33087	5256	27625536
Desember 2025	41208	35185	6023	36276529

$$\begin{aligned}
 \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \\
 &= \sqrt{\frac{132496 + 872356 + 318096 + 129600 + 7661824 + 8346321 + \dots + 36276529}{22}} \\
 &= \sqrt{\frac{227899911}{22}} \\
 &= 3218.52
 \end{aligned}$$

Berdasarkan Tabel 4.13, nilai kuadrat kesalahan (*Squared Error*) pada setiap periode pengujian mengalami peningkatan seiring bertambahnya selisih antara data aktual dan hasil prediksi. Pada awal periode pengujian, seperti Maret

hingga Juni 2024, nilai kesalahan masih relatif kecil dengan nilai *Squared Error* berkisar antara 129.600 hingga 872.356. Namun, memasuki periode Oktober hingga Desember 2025, nilai kesalahan meningkat cukup signifikan hingga mencapai 36.276.529 pada Desember 2025. Kondisi tersebut menunjukkan bahwa model mengalami kesulitan mengikuti peningkatan jumlah penumpang pada beberapa periode akhir pengujian. Meskipun demikian, sebagian besar hasil prediksi masih mampu mengikuti arah perubahan data aktual.

Hasil evaluasi menghasilkan nilai RMSE sebesar 3.218,52. Nilai tersebut menunjukkan bahwa rata-rata penyimpangan hasil prediksi terhadap data aktual berada pada kisaran 3.218 penumpang untuk setiap periode pengujian. Jika dibandingkan dengan jumlah penumpang aktual yang berkisar antara 32 ribu hingga 41 ribu penumpang, tingkat kesalahan tersebut masih tergolong rendah. Hal ini menunjukkan bahwa model tetap mampu memberikan hasil prediksi yang cukup baik meskipun terdapat beberapa periode dengan penyimpangan yang lebih besar.



Gambar 4.4 Grafik Hasil Peramalan Rasio 70:30

Gambar 4.4 menunjukkan hasil peramalan jumlah penumpang kereta api di Pulau Jawa dengan rasio 70:30. Berdasarkan grafik tersebut, terlihat bahwa model

mampu mengikuti pola umum pergerakan data aktual selama periode pengujian. Namun, pada beberapa periode terdapat selisih yang lebih besar antara hasil prediksi dan data aktual, terutama ketika terjadi perubahan jumlah penumpang yang cukup tajam. Kondisi ini menunjukkan bahwa model masih dapat menangkap tren data, tetapi belum mampu mengikuti seluruh fluktuasi yang terjadi secara optimal. Akibatnya, tingkat kesalahan prediksi pada beberapa periode menjadi lebih besar dibandingkan skenario pengujian lainnya.

Meskipun demikian, hasil evaluasi menunjukkan bahwa model tetap memiliki kinerja yang baik dengan nilai MAPE sebesar 7,04% dan RMSE sebesar 3.218,52. Nilai MAPE yang berada di bawah 10% menandakan bahwa tingkat akurasi peramalan masih termasuk dalam kategori sangat baik. Namun, jika dibandingkan dengan skenario pembagian data 80:20 dan 90:10, rasio 70:30 menghasilkan nilai MAPE dan RMSE yang lebih tinggi sehingga tingkat ketepatan prediksinya menjadi lebih rendah. Kondisi ini diduga terjadi karena jumlah data pelatihan yang lebih sedikit membuat model memperoleh informasi historis yang terbatas untuk mempelajari pola tren, pola musiman, serta pengaruh variabel eksogen. Akibatnya, meskipun model masih mampu memberikan hasil peramalan yang baik, performanya belum dapat menyamai hasil yang diperoleh pada skenario pembagian data 80:20.

4.6 Pembahasan

Hasil analisis menunjukkan bahwa model SARIMAX (2,1,2)(1,0,0) merupakan model yang paling sesuai untuk digunakan dalam peramalan jumlah penumpang kereta api di Pulau Jawa. Pemilihan model ini dilakukan melalui beberapa tahapan, mulai dari pengujian stasioneritas, identifikasi orde model, estimasi parameter, uji kesesuaian model, hingga evaluasi performa peramalan. Dari rangkaian proses tersebut, model yang terpilih mampu menggambarkan karakteristik data dengan cukup baik karena dapat menangkap pola tren dan musiman yang terdapat pada data penumpang. Selain itu, model juga mengakomodasi pengaruh variabel eksternal yang digunakan dalam penelitian, yaitu Hari Raya Idul Fitri, libur antar semester, serta periode Natal dan Tahun Baru. Hasil ini menunjukkan bahwa faktor kalender turut berkontribusi terhadap perubahan jumlah penumpang kereta api, sehingga keberadaannya perlu dipertimbangkan dalam proses peramalan.

Karakteristik data menunjukkan adanya pola tren dan musiman yang cukup kuat. Hal ini terlihat dari perlunya proses *non-seasonal differencing* hingga orde kedua sebelum data mencapai kondisi stasioner. Kebutuhan proses *differencing* hingga orde kedua menunjukkan bahwa data penumpang kereta api memiliki tingkat ketidakstasioneran yang cukup tinggi. Kondisi tersebut sejalan dengan perubahan pola mobilitas masyarakat selama periode pandemi COVID-19 dan masa pemulihan setelahnya. Pada periode tersebut terjadi pembatasan perjalanan dan larangan mudik yang menyebabkan jumlah penumpang mengalami penurunan

drastis dibandingkan tahun berikutnya. Setelah pembatasan dicabut, jumlah penumpang kembali meningkat sehingga membentuk pola tren yang kuat pada data.

Di samping adanya pengaruh tren yang terbentuk akibat perubahan kondisi selama dan setelah pandemi, pergerakan jumlah penumpang kereta api juga dipengaruhi oleh faktor eksternal. Hasil penelitian menunjukkan bahwa variabel hari libur yang digunakan dalam model memberikan kontribusi yang signifikan dalam menjelaskan perubahan jumlah penumpang. Kondisi ini menunjukkan bahwa peningkatan mobilitas masyarakat pada saat Idul Fitri, libur antar semester, serta Natal dan Tahun Baru menjadi salah satu faktor yang menyebabkan terjadinya lonjakan penumpang pada periode tertentu.

Berdasarkan temuan tersebut, penggunaan model SARIMAX menjadi relevan karena mampu menggabungkan komponen deret waktu dengan variabel eksogen dalam satu kerangka pemodelan. Hasil penelitian ini juga sejalan dengan penelitian Hayati et al. (2021) yang menunjukkan bahwa penambahan variabel efek kalender pada model SARIMAX dapat meningkatkan kemampuan model dalam menangkap perubahan jumlah penumpang yang dipengaruhi oleh hari raya dan periode libur tertentu.

Sejalan dengan hasil tersebut, model SARIMAX yang dibangun pada penelitian ini juga menunjukkan kinerja yang baik berdasarkan hasil pengujian. Seluruh parameter model memiliki nilai *p-value* kurang dari 0,05 sehingga signifikan secara statistic. Selain itu, hasil uji diagnostik menggunakan uji *Ljung-Box* menghasilkan nilai *p-value* sebesar 0,113359 pada lag 12, 0,600111 pada lag 24, dan 0,915127 pada lag 36. Seluruh nilai tersebut lebih besar dari 0,05 sehingga

residual model tidak mengandung autokorelasi yang signifikan dan telah memenuhi asumsi *white noise*. Dengan demikian, model mampu menjelaskan pola data dengan baik sehingga sisa kesalahan yang dihasilkan bersifat acak.

Berdasarkan hasil pengujian, model SARIMAX (2,1,2)(1,0,0) dipilih sebagai model terbaik dalam penelitian ini. Pemilihan model didasarkan pada beberapa kriteria, yaitu seluruh parameter model memiliki nilai *p-value* kurang dari 0,05 sehingga signifikan secara statistik, residual telah memenuhi asumsi *white noise* berdasarkan uji *Ljung-Box*, serta model menghasilkan tingkat peramalan yang sangat baik.

Tabel 4.14 Hasil Skenario Pengujian

Model	Rasio pembagian data	MAPE	RMSE
SARIMAX (2,1,2)(1,0,0)	90:10	5,05%	2407,38
	80:20	3,75%	1887,12
	70:30	7,04%	3218,52

Berdasarkan Tabel 4.14, model SARIMAX (2,1,2)(1,0,0) diuji menggunakan tiga skenario pembagian data, yaitu rasio 90:10, 80:20, dan 70:30. Hasil pengujian menunjukkan bahwa setiap skenario menghasilkan tingkat akurasi yang berbeda meskipun menggunakan model yang sama. Perbedaan tersebut terlihat dari nilai MAPE dan RMSE yang diperoleh pada masing-masing skenario pengujian. Semakin kecil nilai MAPE dan RMSE, maka semakin baik kemampuan model dalam menghasilkan prediksi yang mendekati data aktual. Oleh karena itu, kedua ukuran tersebut digunakan sebagai dasar untuk membandingkan performa model pada setiap rasio pembagian data.

Berdasarkan hasil pengujian, rasio pembagian data 80:20 menghasilkan nilai MAPE sebesar 3,75% dan RMSE sebesar 1.887,12, yang merupakan nilai

terendah dibandingkan dua skenario lainnya. Sementara itu, rasio 90:10 menghasilkan MAPE sebesar 5,05% dengan RMSE sebesar 2.407,38, sedangkan rasio 70:30 menghasilkan MAPE sebesar 7,04% dan RMSE sebesar 3.218,52. Seluruh nilai MAPE masih berada di bawah 10%, sehingga ketiga skenario termasuk dalam kategori peramalan sangat akurat. Meskipun demikian, rasio 80:20 menunjukkan tingkat kesalahan prediksi yang paling kecil berdasarkan kedua metrik evaluasi tersebut. Hasil ini menunjukkan bahwa model mampu memberikan performa terbaik ketika menggunakan pembagian data 80:20.

Perbedaan hasil pada setiap skenario menunjukkan bahwa proporsi data pelatihan dan data pengujian memengaruhi kinerja model dalam melakukan peramalan. Pada rasio 70:30, jumlah data pelatihan yang lebih sedikit menyebabkan informasi yang digunakan model untuk mempelajari pola historis menjadi lebih terbatas sehingga tingkat kesalahan prediksi meningkat. Sebaliknya, pada rasio 90:10, jumlah data pelatihan memang lebih banyak, tetapi data pengujian yang lebih sedikit menyebabkan hasil evaluasi tidak seoptimal rasio 80:20. Dengan demikian, rasio 80:20 memberikan keseimbangan yang lebih baik antara proses pembelajaran model dan proses evaluasi sehingga menghasilkan tingkat akurasi yang lebih tinggi. Berdasarkan hasil tersebut, rasio pembagian data 80:20 dipilih sebagai konfigurasi terbaik dan digunakan pada tahap peramalan volume penumpang kereta api di Pulau Jawa.

4.7 Integrasi Islam

Penelitian ini dilakukan untuk menghasilkan peramalan jumlah penumpang kereta api di Pulau Jawa menggunakan model SARIMAX. Hasil peramalan yang diperoleh diharapkan dapat digunakan sebagai dasar dalam perencanaan operasional transportasi kereta api secara lebih efektif. Informasi mengenai jumlah penumpang pada periode mendatang dapat membantu pihak pengelola dalam menyesuaikan kapasitas layanan, jadwal perjalanan, serta alokasi sumber daya yang dibutuhkan.

Pentingnya pengelolaan sumber daya secara efisien juga sejalan dengan ajaran Islam sebagaimana dijelaskan dalam QS. Al-Isra' ayat 26 yang melarang perilaku boros dan berlebihan dalam penggunaan sumber daya.

وَأْتِ ذَا الْقُرْبَىٰ حَقَّهُ وَالْمِسْكِينَ وَابْنَ السَّبِيلِ وَلَا تُبَذِّرْ تَبْذِيرًا

“Dan berikanlah haknya kepada kerabat dekat, juga kepada orang miskin dan orang yang dalam perjalanan; dan janganlah kamu menghambur-hamburkan (hartamu) secara boros.”

Menurut Kementerian Agama Republik Indonesia (2010), QS. Al-Isra' ayat 26 mengajarkan agar manusia menggunakan harta dan sumber daya secara bijaksana serta menghindari pemborosan. Hal tersebut dapat diterapkan dalam pengelolaan transportasi kereta api, di mana penyediaan kapasitas layanan perlu disesuaikan dengan kebutuhan yang sebenarnya. Apabila jumlah penumpang diperkirakan terlalu rendah tetapi kapasitas yang disediakan terlalu besar, maka dapat menjadi pemborosa sumber daya operasional. Selain mengajarkan pengelolaan sumber daya secara bijaksana, QS. Al-Isra' ayat 26 juga menunjukkan

perhatian Islam terhadap kebutuhan orang yang sedang melakukan perjalanan. Menurut Kementrian Agama Republik Indonesia (2010) orang yang sedang dalam perjalanan membutuhkan bantuan agar dapat melanjutkan perjalanannya dengan baik. Nilai tersebut sejalan dengan pentingnya penyediaan layanan transportasi yang mampu mendukung mobilitas masyarakat secara optimal. Dengan adanya perencanaan yang baik, layanan transportasi dapat diberikan secara lebih tepat sehingga kebutuhan masyarakat yang melakukan perjalanan dapat terpenuhi dengan baik.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil identifikasi dan pengujian yang telah dilakukan, model SARIMAX (2,1,2)(1,0,0) dipilih sebagai model terbaik untuk meramalkan volume penumpang kereta api di Pulau Jawa. Model ini telah memenuhi uji signifikansi parameter karena seluruh parameter memiliki nilai *p-value* di bawah 0,05. Selain itu, hasil uji diagnostik residual menggunakan uji Ljung-Box menunjukkan bahwa residual model telah memenuhi asumsi *white noise*. Variabel eksogen yang terdiri atas Hari Raya Idul Fitri, libur antar semester, serta Natal dan Tahun Baru juga terbukti membantu model dalam menangkap perubahan jumlah penumpang yang dipengaruhi oleh faktor kalender. Dari tiga skenario pembagian data yang diuji, rasio 80:20 memberikan hasil evaluasi terbaik dengan nilai MAPE sebesar 3,75% dan RMSE sebesar 1.887,12. Hasil tersebut menunjukkan bahwa model memiliki tingkat akurasi yang sangat baik dalam melakukan peramalan volume penumpang kereta api di Pulau Jawa.

5.2 Saran

Berdasarkan proses dan hasil penelitian yang telah diperoleh, penulis memberikan beberapa saran yang diharapkan dapat menjadi bahan pertimbangan bagi peneliti selanjutnya, yaitu sebagai berikut.

1. Penelitian selanjutnya disarankan menggunakan data historis dengan periode yang lebih panjang agar model dapat mempelajari pola tren, pola musiman, serta pengaruh variabel eksogen secara lebih optimal.
2. Penelitian berikutnya juga dapat menambahkan variabel lain yang berpotensi memengaruhi jumlah penumpang, seperti kondisi pandemi atau variabel intervensi lainnya, sehingga model mampu menjelaskan pola data dengan lebih baik.
3. Selain itu, penelitian selanjutnya dapat membandingkan performa model SARIMAX dengan metode peramalan lainnya. Perbandingan tersebut diharapkan dapat memberikan gambaran mengenai metode yang menghasilkan tingkat akurasi terbaik dalam peramalan volume penumpang kereta api.

DAFTAR PUSTAKA

- Ardesfira, G., Zedha, H. F., Fazana, I., Rahmadhiyanti, J., Rahima, S., & Anwar, S. (2022). PERAMALAN NILAI TUKAR RUPIAH TERHADAP DOLLAR AMERIKA DENGAN MENGGUNAKAN METODE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA). *Jambura Journal of Probability and Statistics*, 3(2), 71–84. <https://doi.org/10.34312/jjps.v3i2.15469>
- Badan Pusat Statistik. (2025a). *Berita Resmi Statistik No. 50/06/Th. XXVIII: Perkembangan Transportasi Nasional April 2025*.
- Badan Pusat Statistik. (2025b). *Penduduk, Laju Pertumbuhan Penduduk, Distribusi Persentase Penduduk, Kepadatan Penduduk, Rasio Jenis Kelamin Penduduk Menurut Provinsi, 2025*.
- Badan Pusat Statistik. (2026). *Data Penumpang KAI 2023*.
- Cania, S., Mailisa, D., *2, P., Rianjaya, I. D., Sains, F., Uin, T., & Padang, I. B. (2023). Penerapan Model Seasonal Autoregressive Integrated Moving Average (SARIMA) pada Jumlah Penumpang Kereta Api di Sumatera Barat. In *Journal of Science and Technology* (Vol. 3, Number 2).
- Fungki Wahyu, & Billy Hendrik. (2023). Perbandingan Algoritma Time Series Dan Fuzzy Inference System Dalam Analisis Data Deret Waktu. *Jurnal Penelitian Teknologi Informasi Dan Sains*, 1(3), 16–24. <https://doi.org/10.54066/jptis.v1i3.711>
- Garini, F. C., & Anbiya, W. (2022). Application of GARCH Forecasting Method in Predicting The Number of Rail Passengers (Thousands of People) in Jabodetabek Region. *Jurnal Matematika, Statistika Dan Komputasi*, 18(2), 198–223. <https://doi.org/10.20956/j.v18i2.18382>
- Hayati, Debataraja, N. N., & Martha, S. (2021). *PREDIKSI DATA JUMLAH PENUMPANG KERETA DENGAN EFEK VARIASI KALENDER PADA MODEL SARIMAX* Hayati, Naomi Nussyana Debataraja, Shantika Martha INTISARI.
- Heru Widiyanto, M., & Mayasari, R. (2023). IMPLEMENTASI TIME SERIES PADA DATA PENJUALAN DI GAIKINDO MENGGUNAKAN ALGORITMA SEASONAL ARIMA. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Number 3). <https://files.gaikindo.or.id/>
- Hyndman, R. J., & Athanasopoulos George. (2021). *Forecasting: Principles and practice (3rd ed.)*. OTexts.

- Juliarto, H. K., Purnamasari, I., & Prangga, S. (2024). PERAMALAN PEREDARAN UANG KARTAL DI INDONESIA MENGGUNAKAN MODEL HYBRID SARIMAX-NEURAL NETWORK. *Jurnal Gaussian*, 12(4), 465–476. <https://doi.org/10.14710/j.gauss.12.4.465-476>
- Kementrian Agama Republik Indonesia. (2010). *Mukadimah al-Qur'an dan tafsirnya*. Kementerian Agama Republik Indonesia.
- Khaw, B., Irwanto, R., Yunis, R., & Elly, E. (2025). Analisis Time Series dan Perancangan Dashboard untuk Memprediksi Penjualan dengan Metode Prophet dan SARIMAX. *Jurnal Sifo Mikroskil*, 26(2). <https://doi.org/10.55601/jsm.v26i2.1797>
- Kurnia Hidayat, A., & Wijayanto, A. W. (2025). *Peramalan Volume Timbulan Sampah dengan Memanfaatkan Indeks Google Trends (Anang Kurnia Hidayat)*. <https://doi.org/https://doi.org/10.34123/semnasoffstat.v2025i1.2514>
- Lifa, L., & Yuliansyah, H. (2025). Peramalan Jumlah Pengunjung Wisata Edukasi Museum Menggunakan Kombinasi Moving Average Dan Model Prophet. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(7), 2033–2054. <https://doi.org/10.52436/1.jpti.898>
- Mahmudi, Amalia Mehri, H., & Inna, S. (2024). PENGARUH VARIABEL EKSOGEN TERHADAP PEMODELAN CURAH HUJAN DI KOTA BANDUNG. *Jurnal Pendidikan Matematika Dan Matematika*.
- Mendila, S. A., Utami, I. T., & Kartikasari, P. (2023). PERAMALAN JUMLAH PENUMPANG KERETA API DI PULAU JAWA MENGGUNAKAN METODE HOLT WINTERS EXPONENTIAL SMOOTHING DAN FUZZY TIME SERIES MARKOV CHAIN. *Jurnal Gaussian*, 12(1), 104–115. <https://doi.org/10.14710/j.gauss.12.1.104-115>
- Mirdan Faris, R., Kurniaji, K., Budiman, D., Wijaya Kusuma, M., & Lestari, F. (2024). Analisis Peramalan Produksi Tanaman Kelapa Sawit Menggunakan Metode Arima pada PTPN Kebun Sukamaju Article Info ABSTRAK. In *Jurnal Bisnis dan Manajemen West Science* (Vol. 3, Number 03).
- Pakaya, D. N. P., Achmad, N., Hasan, I. K., Wungguli, D., & Abdussamad, S. N. (2025). Prediksi Wisatawan Mancanegara di Indonesia Menggunakan Metode SARIMAX dengan Efek Variasi Kalender Libur Nasional. *Jurnal Riset Mahasiswa Matematika*, 4(6), 287–300. <https://doi.org/10.18860/jrmm.v4i6.34937>
- Syam, A. R. P. (2022). Application of the Autoregressive Integrated Moving Average Exogenous (ARIMAX) with Calendar Variation Effect Method for Forecasting Chocolate Data in Indonesia and the United

States. *Jurnal Matematika, Statistika Dan Komputasi*, 18(2), 224–236.
<https://doi.org/10.20956/j.v18i2.18460>

Tsabita, R. H., Susetyoko, R., & Satriyanto, E. (2025). *Perbandingan Kinerja ARIMAX dan Fuzzy Time Series Multi Factor pada Peramalan Data Nilai Tukar USD*.

Utomo, P., & Fanani, A. (2020). Peramalan Jumlah Penumpang Kereta Api di Indonesia Menggunakan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA). In *Jurnal Mahasiswa Matematika ALGEBRA* (Vol. 1, Number 1).

Yunianto, R. M., & Muklason, A. (2025). Penerapan Metode Hybrid LSTM dan Seasonal Decomposition untuk Peramalan Permintaan Tiket Harian PT KAI. *ILKOMNIKA: Journal of Computer Science and Applied Informatics E*, 7(2), 168–181.
<https://doi.org/10.28926/ilkomnika.v7i2.733>