

SISTEM PREDIKSI VOLUME PENUMPANG HARIAN KRL YOGYAKARTA-SOLO MENGUNAKAN MODEL HYBRID SARIMAX-PROPHET

Muhammad Ilham 'Aziiz Alfarobi^{1*}, Nurmalitasari², Ratna Puspita Indah³

^{1,2,3} Universitas Duta Bangsa Surakarta, Indonesia

^{1*}220101027@mhs.udb.ac.id, ²nurmalitasari@udb.ac.id, ³ratna_puspita@udb.ac.id

DOI: <https://doi.org/10.46880/mtk.v12i2.5721>

ABSTRACT

The operation of the Yogyakarta-Solo Commuter Line (KRL) since 2021 has become the backbone of transportation in the Yogyakarta Special Region and Central Java. However, highly dynamic fluctuations in passenger volume pose challenges for operational optimization. This research aims to develop an accurate daily passenger volume prediction system with a 30 day forecasting horizon to mitigate overcrowding and fleet inefficiency. The methodology employed is CRISP-DM, proposing a layered hybrid architecture based on residual modeling. In this model, SARIMAX serves as the primary pattern modeler (Layer 1), while Facebook Prophet acts as a residual corrector (Layer 2), optimized with selective correction mechanisms and daily adaptive weights. The research data covers the period from January 2025 to January 2026, totaling 396 observations. The test results show that the hybrid model provides the best performance compared to single models, achieving a Mean Absolute Percentage Error (MAPE) of 9.66% and a Mean Absolute Error (MAE) of 2,739 passengers per day. Utilizing historical data from January 2025 to January 2026, the modeling results are integrated into an interactive Streamlit dashboard as a practical decision support tool for KAI Commuter's proactive operational planning.

Keywords: *Hybrid Model, Passenger Volume Prediction, SARIMAX-Prophet, Yogyakarta-Solo Commuter Line.*

I. PENDAHULUAN

Pengoperasian Kereta Rel Listrik (KRL) Yogyakarta-Solo sejak awal 2021, telah berkembang menjadi tulang punggung transportasi di kawasan metropolitan Jawa Tengah dan Daerah Istimewa Yogyakarta (DIY) [1]. Layanan ini menjadi pilihan utama berbagai kelompok masyarakat mulai dari pelajar harian, mahasiswa, hingga wisatawan karena menawarkan tarif yang sangat kompetitif sebesar Rp8.000 didukung dengan waktu tempuh yang efisien, bebas macet, dan ramah lingkungan [2], [3]. Daya tarik komprehensif tersebut memicu ledakan pertumbuhan mobilitas yang masif sejak awal peluncurannya, jika pada tahun pertama beroperasi (2021) KRL ini melayani 1.739.454 pengguna, PT KCI mencatat total volume tahunan tersebut meroket hingga mencapai 9.015.768 orang pada tahun 2025, dengan peningkatan sebesar 13 persen dari tahun 2024 [4]. Keberagaman pengguna dan tingginya volume tersebut pada akhirnya memicu tren pergerakan harian penumpang yang sangat dinamis dari waktu ke waktu.

Dinamika pergerakan penumpang tersebut menghadirkan tantangan operasional yang kompleks bagi manajemen, terutama karena tingginya sensitivitas terhadap faktor musiman [5]. Lonjakan volume tidak hanya terpusat pada jam sibuk harian, tetapi juga meningkat secara signifikan pada akhir pekan akibat pergerakan wisatawan, serta mengalami fluktuasi ekstrem pada momentum hari libur nasional dan musim liburan panjang [6]. Variabilitas yang tinggi ini berdampak langsung pada efisiensi layanan, khususnya dalam menyulitkan optimalisasi jadwal perjalanan, pengaturan stamformasi (jumlah gerbong), hingga penyiapan sarana prasarana pendukung di stasiun [7]. Apabila fluktuasi ini tidak terantisipasi, tingginya risiko penumpukan penumpang (*overcrowding*) akan secara langsung menurunkan tingkat kenyamanan dan kualitas pelayanan [8]. Oleh karena itu, diperlukan sebuah

sistem peramalan deret waktu yang akurat untuk mengantisipasi ketidakpastian beban operasional tersebut.

Berbagai metode peramalan deret waktu telah banyak diaplikasikan untuk memprediksi dinamika transportasi publik [9]. Di antara metode tersebut, model *AutoRegressive Integrated Moving Average* (ARIMA) merupakan pendekatan klasik yang memanfaatkan komponen *autoregressive* dan *moving average* untuk menghasilkan prediksi akurat setelah melalui proses diferensiasi data non-stasioner [10]. Sementara itu, *Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors* (SARIMAX) Sebagai variasi ARIMA yang mempertimbangkan efek musiman dan variabel eksternal [11], model ini dapat mengintegrasikan parameter musiman tambahan untuk menangkap pola berulang secara periodik pada deret waktu [12]. Meskipun kedua model klasik ini sangat tangguh dalam memodelkan struktur linier dan tren dasar, performanya sering kali menurun secara signifikan dan menghasilkan galat (*error*) yang besar ketika dihadapkan pada volatilitas ekstrem serta anomali non-linier seperti momentum hari libur nasional [13].

Sebagai alternatif untuk mengatasi keterbatasan tersebut, pendekatan berbasis machine learning seperti Facebook Prophet telah terbukti tangguh dalam mengomodasi data yang memiliki efek hari libur (*holiday effects*) secara dinamis [14]. Dalam domain transportasi, efektivitas model Prophet telah diuji oleh Prakoso dalam memproyeksikan volume penumpang Trans Metro Bandung [15], serta oleh Yudistira untuk mengestimasi pergerakan penumpang angkutan udara [16]. Kendati demikian, karakteristik Prophet yang memodelkan deret waktu sebagai kurva aditif cenderung mengabaikan struktur autokorelasi mendasar yang terdapat pada sisa galat harian [17]. Akibatnya, tingkat presisi prediksi harian dari model tunggal ini

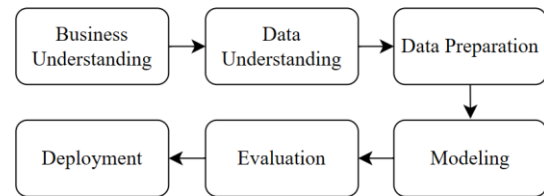
sering kali kurang konsisten dibandingkan dengan model statistik murni, sehingga keterbatasan pada masing-masing pendekatan tunggal tersebut membuka ruang untuk optimasi lebih lanjut melalui pemodelan hybrid.

Untuk mengatasi keterbatasan dari masing-masing model tunggal tersebut, penelitian ini mengusulkan pengembangan sistem peramalan volume penumpang harian menggunakan arsitektur Hybrid SARIMAX-Prophet. Melalui pendekatan integratif ini, model SARIMAX diimplementasikan sebagai Layer 1 yang bertugas menyerap komponen tren linier, musiman dasar mingguan, serta pengaruh variabel eksogen secara statistik [18]. Sisa galat (*residuals*) yang tidak terserap pada lapisan pertama kemudian diekstrak dan diumpungkan ke dalam algoritma Facebook Prophet sebagai Layer 2 untuk dikoreksi menggunakan regressor awal pekan dan mekanisme bobot adaptif harian. Pendekatan kombinasi berjenjang ini diproyeksikan mampu menghasilkan estimasi yang lebih kokoh dan signifikan dibandingkan performa masing-masing model secara individual dalam menghadapi fluktuasi data transportasi yang kompleks [19].

Integrasi pemodelan hybrid ini pada akhirnya tidak hanya ditujukan untuk mencapai akurasi metrik evaluasi yang tinggi di atas kertas, melainkan ditransformasikan ke dalam bentuk dashboard interaktif berbasis web menggunakan framework Streamlit sebagai langkah hilirisasi operasional. Pemanfaatan Streamlit sebagai media visualisasi taktis dalam riset prediktif sejalan dengan penelitian Lisnawati yang sukses mengimplementasikan framework tersebut untuk sistem prediksi kunjungan wisatawan [20]. Implementasi sistem yang dikembangkan dalam penelitian ini secara praktis akan berfungsi sebagai alat bantu keputusan (*Decision Support System*) harian bagi manajemen perusahaan [21]. Secara akademis, penelitian ini bertujuan untuk merancang sistem peramalan yang mampu memberikan estimasi volume penumpang untuk rentang waktu 7 hingga 30 hari ke depan dengan target batas toleransi galat Mean Absolute Percentage Error (MAPE) di bawah 10% agar memenuhi kriteria industri sebagai peramalan yang akurat. Berdasarkan tujuan tersebut, pertanyaan riset yang diajukan dalam penelitian ini adalah: (1) Bagaimana efektivitas implementasi model Hybrid SARIMAX-Prophet dalam memprediksi volume penumpang harian KRL Yogyakarta-Solo? dan (2) Model manakah yang memberikan performa akurasi terbaik di antara model tunggal dan model hybrid pada dataset tersebut?

II. METODE

Metodologi penelitian menggunakan Cross-Industry Standard Process for Data Mining (CRISP DM) yang terdiri dari enam tahapan utama: Pemahaman Bisnis (Business Understanding), Pemahaman data (Data Understanding), Pengolahan Data (Data Preparation), Pemodelan (Modeling), Evaluasi (Evaluation), Penyebaran (Deployment) [22].



Gambar 1. Alur Metodologi

Sumber Data

Metodologi penelitian ini menggunakan data sekunder volume penumpang harian KRL Yogyakarta-Solo periode 1 Januari 2025 hingga 31 Januari 2026 (396 observasi) dari PT KCI, yang dibagi secara kronologis menjadi data latih (80%) dan data uji (20%) setelah melalui praproses interpolasi linear untuk menghaluskan distorsi anomali ekstrem Lebaran.

Business Understanding

Pada tahap awal ini, dilakukan identifikasi terhadap permasalahan manajerial yang dihadapi oleh KAI Commuter. Masalah utama yang ditemukan adalah fluktuasi volume penumpang yang sangat dinamis, khususnya pada periode hari libur nasional, hari raya, dan akhir pekan, yang memicu ketidakpastian dalam pengelolaan beban operasional. Dari pemahaman ini dirumuskan dua pertanyaan penelitian. Pertama, bagaimana efektivitas implementasi model Hybrid SARIMAX-Prophet dalam memprediksi volume penumpang harian KRL Yogyakarta-Solo. Kedua, model manakah yang memberikan performa akurasi terbaik di antara model tunggal dan model hybrid pada dataset tersebut. Tujuan akhir dari fase ini adalah merancang sistem peramalan yang mampu memberikan estimasi volume penumpang untuk rentang waktu 7 hingga 30 hari ke depan dengan target batas toleransi galat *Mean Absolute Percentage Error* (MAPE) di bawah 10% agar memenuhi kriteria industri sebagai peramalan yang sangat akurat (*highly accurate forecasting*). Melalui proyeksi tersebut, pihak pengelola diharapkan dapat mengambil keputusan operasional secara proaktif guna memitigasi risiko penumpukan penumpang dan mengoptimalkan efisiensi armada.

Data Understanding

Data yang digunakan dalam penelitian ini adalah volume penumpang harian KRL Yogyakarta-Solo yang diperoleh dari Website PT Kereta Commuter Indonesia (KAI Commuter) [23]. Dataset mencakup periode pengamatan dari 1 Januari 2025 hingga 31 Januari 2026, yang menghasilkan total 396 observasi harian. Berdasarkan pemeriksaan awal, dataset bersifat lengkap tanpa adanya nilai yang hilang (*missing values*). Berdasarkan Analisis statistik deskriptif menunjukkan bahwa rata-rata volume penumpang harian adalah sebesar 25.840 orang. Fluktuasi data terekam cukup ekstrem dengan nilai minimum 14.888 penumpang pada periode pra-Lebaran dan nilai maksimum mencapai 45.536 penumpang pada periode pasca-Lebaran. Melalui tahapan *Exploratory Data Analysis* (EDA), diidentifikasi tiga karakteristik fundamental yang menjadi landasan dalam pengembangan model. Pertama, terdapat pola mingguan yang sangat kuat dan konsisten di mana

volume penumpang pada akhir pekan (Sabtu dan Minggu) rata-rata 15-16% lebih tinggi dibandingkan rata-rata harian. Sebaliknya, pada hari Selasa dan Rabu, volume menurun sekitar 10% dari rata-rata. Hal ini mengonfirmasi dominasi pergerakan wisatawan pada rute ini. Kedua, terdapat pola bulanan yang berkaitan dengan musim liburan.

Terdapat tren musiman yang signifikan pada bulan Mei hingga Juli dengan peningkatan volume sebesar 20% yang bertepatan dengan masa libur sekolah. Sementara itu, volume terendah tercatat pada bulan Maret yang bertepatan dengan periode awal Ramadan dan persiapan mudik. Ketiga, ditemukan pola anomali unik berupa penurunan tajam sesaat sebelum Idulfitri, diikuti oleh lonjakan ekstrem (puncak tahunan) pada periode H+1 hingga H+11. Temuan ini, khususnya anomali Lebaran, menuntut penanganan khusus dalam pemodelan untuk mencegah distorsi pada estimasi hasil prediksi.

Data Preparation

Praproses data diterapkan secara konsisten pada ketiga model untuk menjamin validitas perbandingan, meliputi tiga tahapan strategis. Pertama, anomali Lebaran (25 Maret - 10 April 2025) ditangani melalui interpolasi linear selama pelatihan untuk mencegah distorsi estimasi tren jangka panjang, sementara efeknya tetap dimodelkan secara eksplisit melalui variabel dummy. Kedua, dibuat tiga variabel dummy granular yaitu hari sebelum lebaran, pada hari puncak lebaran dan terakhir pasca lebaran guna menangkap sifat asimetris dampak Lebaran terhadap volume penumpang. Ketiga, data dibagi secara kronologis dengan rasio 80:20 (316 hari latih, 80 hari uji) untuk menghindari *data leakage* dan memastikan evaluasi model mencerminkan kondisi prediktif nyata pada data deret waktu.

Modeling

Pada tahap ini dijabarkan model yang dipakai dalam penelitian ini yaitu Prophet, SARIMAX dan Hybrid SARIMAX-prophet

Prophet

Facebook Prophet atau Prophet adalah sebuah model peramalan open-source dikembangkan oleh tim Core Data Science Facebook yang dirancang untuk menangani data deret waktu dengan pola musiman yang kuat dan data historis yang panjang [24]. Berikut rumus dasar dari Prophet:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (1)$$

Diketahui

$y(t)$: Nilai aktual pada waktu t

$g(t)$: Komponen tren

$s(t)$: Komponen Musim

$h(t)$: Efek hari libur

ϵ_t : noise atau error

Model Prophet dipilih sebagai karena kemampuannya dalam mengakomodasi pola tren, musim, dan efek hari libur yang sering ditemukan pada data deret waktu [25]. Pada penelitian ini model prophet diatur menggunakan fungsi pertumbuhan linier dengan mode musiman aditif (*additive*). Untuk mencegah overfitting, musiman tahunan dinonaktifkan (*yearly_seasonality = False*). Sebagai penguat, regressor kustom eksternal berupa

variabel *is_early_week* (Senin-Rabu) ditambahkan secara eksplisit untuk menangkap karakteristik asimetri galat di awal pekan.

SARIMAX

SARIMAX merupakan pengembangan dari model ARIMA (Box-Jenkins) yang dirancang khusus untuk menangani data deret waktu bermusim sekaligus mengintegrasikan pengaruh variabel eksogen [26]. Variabel luar dipilih berdasarkan relevansi teoretisnya dalam memperkuat daya jelaskan model terhadap variabel respons. Struktur model dilambangkan dengan notasi $(p, d, q) \times (P, D, Q)_s$. Secara matematis, hubungan antar variabel dalam model SARIMAX didefinisikan melalui persamaan berikut:

$$\Phi_p(B^s)\phi_p(B)\nabla_s^D\nabla^d y_t = \beta X_t + \Theta_q(B^s)\theta_q(B)\epsilon_t \quad (2)$$

Diketahui:

Komponen Non-Musiman (p, d, q)

$\phi_p(B)$: Parameter AutoRegressive

∇^d : Operator Differencing

$\theta_q(B)$: Parameter MovingAverage

Komponen Musiman $(P, D, Q)_s$

$\Phi_p(B^s)$ dan $\Theta_q(B^s)$: Komponen AR dan MA musiman

∇_s^D : Operator Differencing

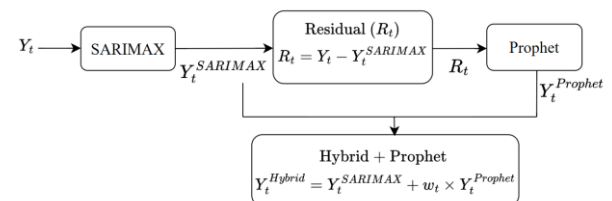
βX_t : Faktor Eksogen

ϵ_t : Error (*white noise*).

Dengan menyertakan variabel eksogen, model SARIMAX tidak hanya bergantung pada pola deret waktu internal, tetapi juga mampu mengolah data kontekstual tambahan untuk hasil prediksi yang lebih komprehensif [27]. Konfigurasi SARIMAX ditentukan melalui pencarian berbasis algoritma Auto ARIMA, menghasilkan orde parameter SARIMAX(2,0,0) × (2,1,1)₇ disertai komponen intercept. Model ini dikonfigurasi untuk menyerap tren linier, musiman mingguan bersiklus harian ($s = 7$), serta variabel eksogen (X) granular yang merepresentasikan variabel dummy libur nasional dan tiga fase libur Lebaran.

Hybrid SARIMAX-Prophet

Model Hybrid SARIMAX-Prophet menggunakan arsitektur sebagai berikut:



Gambar 2. Arsitektur Model Hybrid SARIMAX-Prophet

Penelitian ini mengimplementasikan model Hybrid SARIMAX-Prophet melalui pemodelan berjenjang (*layered modeling*) berbasis sisa galat (*residual-based modeling*). Arsitektur ini menggabungkan ketajaman SARIMAX dalam memodelkan pola dasar dengan fleksibilitas Prophet dalam menyerap sisa variansi non-linear [28]. Formulasi akhir model Hybrid didefinisikan sebagai:

$$Y_t^{Hybrid} = Y_t^{SARIMAX} + w_t \times Y_t^{Prophet} \quad (3)$$

di mana $Y_t^{SARIMAX}$ adalah prediksi dasar Layer 1, w_t adalah bobot adaptif harian, dan $R_t^{Prophet}$ adalah hasil prediksi penyesuaian Layer 2 yang telah dikoreksi.

Pemodelan Dasar (Layer 1 - SARIMAX)

SARIMAX memodelkan deret waktu utama untuk menangkap komponen linear, musiman mingguan, dan efek variabel eksogen kalender. Sisa galat latih diekstrak guna memetakan variansi yang belum dijelaskan oleh Layer 1:

$$R_t = Y_t - Y_t^{SARIMAX} \quad (4)$$

Pencilan ekstrem pada nilai R_t dipangkas (*clipping*) menggunakan batas *Interquartile Range* (IQR) untuk mencegah distorsi pemodelan pada tahap selanjutnya. Penentuan parameter optimal pada model Layer 1 dilakukan melalui tahapan *Auto ARIMA* yang menghasilkan spesifikasi SARIMAX(2,0,0) × (2,1,1)₇ dengan konstanta.

Pemodelan Residu (Layer 2 - Prophet)

Deret waktu galat R_t dimodelkan secara aditif menggunakan Facebook Prophet tanpa komponen tren tahunan. Model diintegrasikan dengan regressor tambahan berupa variabel dummy awal pekan (*is_early_week* untuk hari Senin-Rabu) guna mengakomodasi sebaran galat Layer 1 yang terbukti asimetris. Output dari tahap ini dinotasikan sebagai $R_t^{Prophet}$.

Koreksi Selektif dan Pembobotan Adaptif

Untuk mencegah model Layer 2 memberikan koreksi yang berlebihan (*over-correction*) pada data uji, diterapkan protokol Koreksi Selektif berbasis ambang batas (*threshold*) yang mengacu pada nilai Mean Absolute Error (MAE) acuan dari pengujian model SARIMAX ($MAE_{SARIMAX}$). Jika prediksi intervensi Layer 2 melampaui rata-rata galat Layer 1, nilainya direduksi sebesar 50%. Terakhir, Bobot Adaptif (w_t) diimplementasikan berdasarkan karakteristik hari [29]. Intervensi penyesuaian ditekan pada awal pekan ($w_t = 0.10$ untuk Senin-Rabu) karena Layer 1 terbukti sangat stabil. Sebaliknya, bobot Layer 2 dimaksimalkan pada akhir pekan ($w_t = 0.80 - 0.90$ untuk Kamis-Minggu) guna mengompensasi kecenderungan under-forecasting SARIMAX saat terjadi lonjakan volume penumpang wisata.

Evaluation

Tahap evaluasi bertujuan untuk mengukur performa prediktif ketiga model melalui protokol pengujian yang ketat. Pengujian dilakukan menggunakan metode Cross-Validation (*Rolling Window*) guna memastikan kestabilan model pada berbagai segmen waktu. Metrik yang digunakan meliputi MAE (*Mean Absolute Error*)

dan RMSE (*Root Mean Squared Error*) untuk mengukur besarnya penyimpangan volume penumpang secara absolut [30], serta MAPE dan SMAPE untuk mengukur persentase kesalahan relatif terhadap fluktuasi harian. Coverage untuk mengukur persentase data aktual yang jatuh di dalam rentang Confidence Interval (95%). Hal ini krusial bagi manajemen KAI untuk memahami batasan ketidakpastian (*uncertainty*) dalam skenario terburuk

Deployment

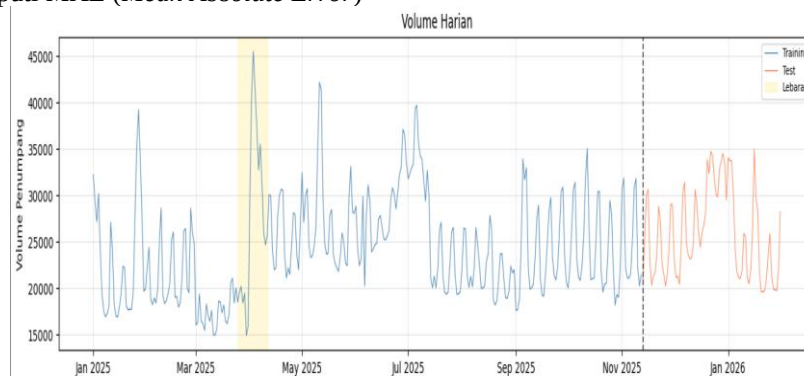
Hasil pemodelan terbaik diintegrasikan ke dalam sebuah Dashboard Interaktif berbasis Streamlit. Tahap ini bertujuan untuk menjembatani kompleksitas algoritma dengan kebutuhan praktis pengambilan keputusan di lapangan. Dashboard interaktif ini mengintegrasikan tiga halaman utama yaitu Dashboard Perbandingan Model yang menampilkan evaluasi komprehensif (metrik, visualisasi prediksi) menggunakan data statis untuk akses cepat. Halaman Prediksi Penumpang yang memungkinkan pengguna memilih model dan horizon prediksi (7/14/30 hari) untuk menghasilkan proyeksi, tabel unduhan, serta rekomendasi operasional otomatis berbasis threshold volume.

Tentang Penelitian yang menyajikan dokumentasi metodologi, dataset, dan konfigurasi model secara transparan sebagai referensi akademis yang dapat diverifikasi. Dengan adanya deployment ini, sistem tidak hanya berhenti sebagai studi teknis, tetapi bertransformasi menjadi alat pendukung keputusan (*Decision Support System*) yang membantu manajemen dalam optimalisasi jumlah petugas stasiun dan alokasi rangkaian gerbong secara proaktif.

III. HASIL DAN PEMBAHASAN

Analisis Karakteristik Data

Analisis awal terhadap volume penumpang KRL Yogyakarta-Solo menunjukkan pola musiman ganda yang sangat kuat. Pola mingguan didominasi oleh lonjakan pada akhir pekan (Sabtu-Minggu) dengan rata-rata kenaikan 15-16% dibandingkan hari kerja, mencerminkan karakteristik rute ini sebagai jalur wisata utama. Secara musiman, lonjakan ekstrem terjadi pada periode pasca-Lebaran dan libur sekolah (Mei-Juli). Temuan anomali asimetris pada masa Lebaran penurunan tajam sebelum hari H dan lonjakan masif setelahnya menjadi justifikasi utama diperlukannya rekayasa fitur berupa variabel dummy granular untuk mencegah distorsi pada model.



Gambar 3. Grafik Time Series Volume Penumpang

Metrik Evaluasi

Tabel 1 menyajikan perbandingan performa antara model tunggal (Prophet dan SARIMAX) dengan model Hybrid yang diusulkan. Model Prophet menunjukkan galat tertinggi MAPE=15,88%, mengindikasikan kesulitan model dalam menangkap fluktuasi harian yang sangat dinamis. SARIMAX memberikan performa yang jauh lebih baik MAPE = 9,75% dengan memanfaatkan variabel eksogen hari libur. Namun, model Hybrid SARIMAX-Prophet berhasil memberikan akurasi terbaik dengan nilai MAPE 9,66% dan MAE 2.739 penumpang per hari. Penurunan nilai RMSE pada model Hybrid juga menunjukkan bahwa model ini lebih stabil dalam menangani fluktuasi besar dibandingkan model lainnya. Model hybrid lebih baik dari model tunggal karena ia mampu menghasilkan kestabilan prediksi pada hari kerja biasa berkat SARIMAX, sekaligus menjaga ketepatan prediksi saat terjadi lonjakan ekstrem di hari libur berkat koreksi Prophet.

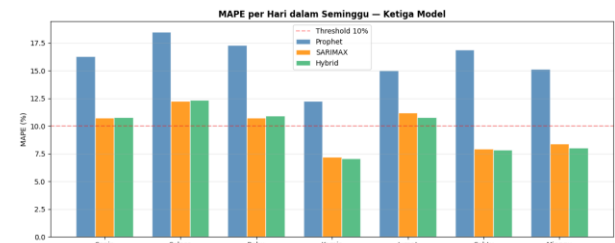
Tabel 1. Hasil Metrik Evaluasi

Metrik	Prophet	SARIMAX	Hybrid
MAE (pnp/hari)	4.492	2.756	2.739
RMSE (pnp/hari)	6.040	4.276	4.283
MAPE(%)	15.88	9.75	9.66
SMAPE(%)	18.16	10.44	10.41
Coverage 95%	78.75	97.50	97.50
CI(%)			
CV MAPE(%)	18.75	7.54	7.20

Analisis MAPE per hari

Untuk memahami di mana model Hybrid memberikan kontribusi terbesar, dilakukan analisis galat per hari. Grafik menunjukkan bahwa SARIMAX sangat

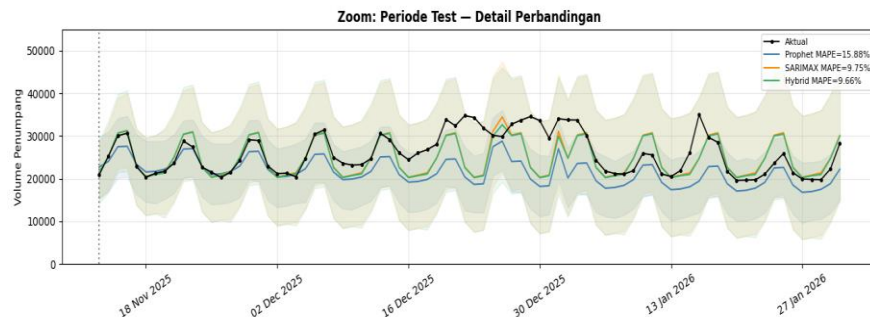
kompetitif pada hari kerja (Senin-Rabu), namun mengalami peningkatan galat pada akhir pekan. Sebaliknya, model Hybrid melalui mekanisme Bobot Adaptif berhasil menekan galat secara signifikan pada hari Kamis hingga Minggu. Hal ini membuktikan bahwa penambahan Layer 2 (Prophet) efektif dalam mengoreksi kecenderungan *under-forecasting* pada periode puncak mobilitas wisatawan, yang sering gagal ditangkap secara sempurna oleh model linear SARIMAX.



Gambar 4. Grafik MAPE per Hari

Visualisasi Prediksi vs Aktual

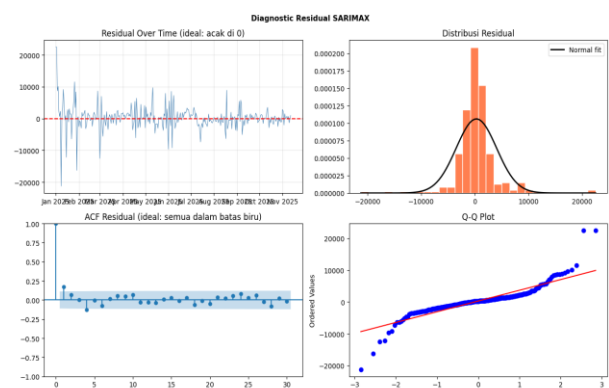
Gambar 5 menunjukkan perbandingan visual antara data aktual dan hasil prediksi ketiga model pada periode data uji. Secara visual, garis prediksi Hybrid (hijau) mampu mengikuti fluktuasi data aktual (hitam) lebih rapat dibandingkan model lainnya. Model Hybrid terbukti lebih responsif dalam menangkap puncak (*peaks*) dan lembah (*troughs*) data, terutama pada masa transisi antara hari kerja dan libur nasional. Hal ini menunjukkan bahwa skema pemodelan residu berhasil mengekstraksi informasi berharga yang tertinggal pada model pertama.



Gambar 5. Grafik Periode Test

Diagnostik Residual

Pada uji diagnostik residu menunjukkan bahwa sisa galat telah terdistribusi secara normal dan tidak lagi mengandung autokorelasi yang signifikan berdasarkan plot ACF dan Uji Ljung-Box. Hal ini mengonfirmasi bahwa arsitektur Hybrid yang disusun telah berhasil menyerap hampir seluruh pola informasi yang tersedia dalam dataset.

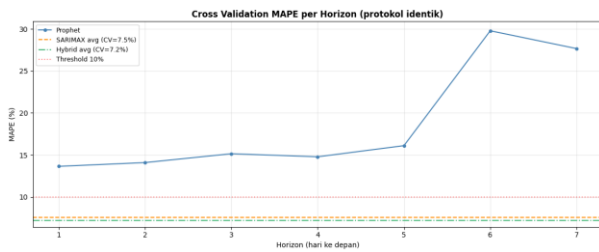


Gambar 6. Grafik Diagnostik Residual

Cross Validation Horizon

Evaluasi melalui Cross-Validation dengan skema 5-fold menunjukkan bahwa model Hybrid memiliki

stabilitas yang baik untuk perencanaan jangka pendek. Pembagian data menjadi 5 bagian yang bergiliran sebagai data uji memastikan bahwa evaluasi model lebih robust dan tidak bias terhadap satu periode tertentu. Galat prediksi tetap berada di bawah ambang batas 10% untuk rentang waktu hingga 7 hari ke depan. Stabilitas ini sangat krusial bagi implementasi praktis, di mana manajemen transportasi memerlukan estimasi yang konsisten untuk merencanakan alokasi petugas dan stamformasi rangkaian kereta setiap minggunya.



Gambar 7. Grafik Cross Validation

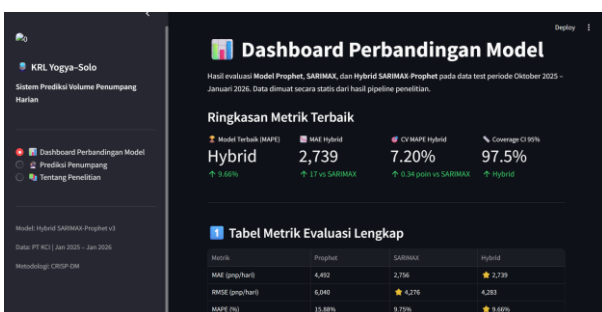
Gambar 7 menunjukkan bahwa model Hybrid SARIMAX-Prophet mencapai nilai CV MAPE terendah sebesar 7,20%, lebih unggul dibandingkan SARIMAX (7,54%) dan Prophet (18,75%).

Deployment Streamlit

Tahap akhir dari penelitian ini adalah pengembangan sistem pemantauan dan prediksi interaktif berbasis web menggunakan framework Streamlit. Implementasi ini bertujuan untuk mentransformasi model Hybrid SARIMAX-Prophet menjadi alat pendukung keputusan (*Decision Support System*) yang dapat diakses secara mudah oleh pihak operasional.

Halaman Perbandingan Model

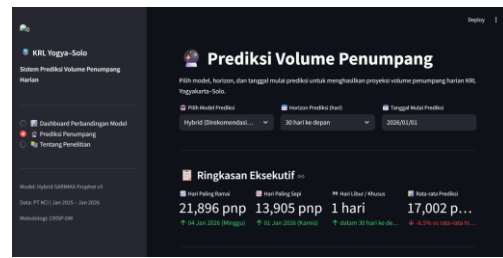
Halaman ini menyajikan komparasi performa antara tiga model yaitu Prophet, SARIMAX, dan Hybrid SARIMAX-Prophet. Fokus utama di sini adalah metrik evaluasi seperti MAE, RMSE, dan MAPE. Grafik perbandingan harian memungkinkan pengguna melihat secara detail di mana model Hybrid berhasil mengoreksi kesalahan dari model tunggal.



Gambar 8. Halaman Perbandingan Model

Halaman Prediksi Volume Penumpang

Halaman Prediksi dirancang untuk kebutuhan proaktif. Pengguna dapat memilih horizon waktu (7 hingga 30 hari ke depan) untuk melihat estimasi volume penumpang. Fitur interaktif seperti slider dan date input memungkinkan perencanaan logistik transportasi dilakukan secara dinamis.



Gambar 9. Halaman Prediksi Volume Penumpang

Halaman Tentang Penelitian

Halaman ini memberikan konteks teknis mengenai metodologi CRISP-DM yang digunakan, arsitektur Hybrid dua lapis (SARIMAX sebagai model utama dan Prophet sebagai korektor residu), serta informasi mengenai dataset yang bersumber dari data resmi PT KCI.



Gambar 10. Halaman Tentang Penelitian

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem prediksi volume penumpang harian KRL Yogyakarta-Solo menggunakan arsitektur Hybrid SARIMAX-Prophet berbasis *residual-based modeling*. Secara empiris, model hybrid ini menghasilkan performa terbaik dalam kategori peramalan sangat akurat dengan nilai MAPE sebesar 9,66% dan MAE 2.739 penumpang per hari untuk horizon prediksi hingga 30 hari ke depan. Keberhasilan ini didorong oleh kemampuan model dalam mengoreksi *under-forecasting* akhir pekan serta penggunaan variabel dummy granular untuk menetralkan guncangan data saat hari raya. Hasil pemodelan ini telah diimplementasikan ke dalam dashboard interaktif Streamlit sebagai *Decision Support System* bagi manajemen KAI Commuter untuk perencanaan operasional secara proaktif. Untuk optimalisasi jangka panjang, direkomendasikan agar pihak manajemen mengintegrasikan dashboard ini ke dalam SOP operasional harian serta melakukan pembaruan parameter model secara berkala menggunakan data terbaru guna menjaga stabilitas akurasi prediksi. Meskipun model Hybrid SARIMAX-Prophet ini berhasil mencapai akurasi yang tinggi, penelitian ini masih memiliki keterbatasan terkait cakupan dataset yang baru berjalan satu tahun, sifat pemodelan yang masih *offline/batch*, serta belum diintegrasikannya faktor eksternal seperti cuaca dan data granular per stasiun. Oleh karena itu, penelitian selanjutnya sangat disarankan untuk memperpanjang rentang waktu historis dataset menjadi 3-5 tahun guna menangkap pola musiman tahunan (*yearly seasonality*) dan tren jangka panjang secara lebih stabil, memperluas fitur eksogen multivariat (seperti kondisi cuaca dan acara besar), mengotomatiskan hyperparameter tuning, serta mengembangkan sistem ke arah online learning yang terintegrasi langsung dengan API transaksi

penumpang secara real-time.

REFERENSI

- [1] F. Fadhilah, M. Y. Hartiman, dan I. Prasetyo, "Analisis Rencana Pengoperasian KRL Lintas Yogyakarta-Kutoarjo Sebagai Pengganti KRDE Prameks".
- [2] C. Noermartanto, H. Risnandar, N. Anggani, dan P. Hmp Uns, *Analisis Peningkatan Komuter Di Perkotaan Solo -Yogya Menggunakan Commuter Line*. 2024.
- [3] T. N. Arifin, S. Sodikin, dan T. Tantin, "ANALISA KINERJA OPERASIONAL KRL COMMUTER LINE JURUSAN YOGYAKARTA-SOLO," *Rang Tek. J.*, vol. 8, no. 2, hlm. 273–283, Jun 2025, doi: 10.31869/rtj.v8i2.6228.
- [4] "KAI Commuter." Diakses: 19 Mei 2026. [Daring]. Tersedia pada: <https://kci.id/informasi-publik/berita/capaian-membanggakan-volume-pengguna-commuter-line-wilayah-6-yogyakarta-tembus-10-juta-sepanjang-2025>
- [5] R. M. Yuniarto dan A. Muklason, "Penerapan Metode Hybrid LSTM dan Seasonal Decomposition untuk Peramalan Permintaan Tiket Harian PT KAI," *ILKOMNIKA*, vol. 7, no. 2, hlm. 168–181, Agu 2025, doi: 10.28926/ilkomnika.v7i2.733.
- [6] "KAI Commuter." Diakses: 19 Mei 2026. [Daring]. Tersedia pada: <https://kci.id/informasi-publik/berita/masa-libur-sekolah-volume-pengguna-krl-yogyakarta-solo-naik-17>
- [7] S. S. S. Sidabutar, S. T. Sitohang, M. J. Samosir, Y. A. Simatupang, dan J. T. Hardinata, "Perbandingan Metode Regresi Linier dan Artificial Neural Network dalam Memprediksi Jumlah Penumpang Kereta Api Nasional," *J. Komput. Teknol. Inf. Sist. Komput. JUKTISI*, vol. 4, no. 3, hlm. 2125–2132, Jan 2026, doi: 10.62712/juktisi.v4i3.822.
- [8] A. C. Damar, D. Safitri, dan S. Sujarwo, "Analisis Dampak Kepadatan Penumpang KRL Terhadap Tingkat Stress Masyarakat Jabodetabek," *J. Ilm. Penelit. Mhs.*, vol. 3, no. 2, hlm. 226–235, Apr 2025, doi: 10.61722/jipm.v3i2.815.
- [9] R. Wati, R. Estriana, R. Praptiwi, B. S. Ilimi, N. R. Oktavian, dan T. Zakiyudin, "Analisis Perbandingan XGBoost, Random Forest, dan LSTM untuk Prediksi Jumlah Penumpang Bus," *J. Inform. Interact. Technol.*, vol. 2, no. 3, hlm. 405–413, Des 2025, doi: 10.63547/jiite.v2i3.73.
- [10] L. D. Jayanti, R. Lestari, F. A. D. Suparno, dan F. Fachruzzaki, "Prediksi Harga Emas Tahun 2024-2025 dengan Metode Autoregressive Integrated Moving Average (ARIMA) pada Aplikasi RStudio," *MARAS J. Penelit. Multidisiplin*, vol. 3, no. 4, hlm. 1275–1289, Okt 2025, doi: 10.60126/maras.v3i4.1253.
- [11] I. K. Chisan dan A. W. Wijayanto, "Forecasting Jumlah Penumpang Pesawat Yogyakarta International Airport dengan Big Data Google Trends dan Variabel Makroekonomi untuk Mendukung Official Statistics," *Semin. Nas. Off. Stat.*, vol. 2024, no. 1, hlm. 615–624, Nov 2024, doi: 10.34123/semnasoffstat.v2024i1.2123.
- [12] L. L. Ibrahim dan E. Kurniati, "Peramalan Jumlah Penumpang Kereta Api Eksekutif di Pulau Jawa Menggunakan Model SARIMA," *J. Ris. Mat.*, hlm. 73–82, Jul 2023, doi: 10.29313/jrm.v3i1.1747.
- [13] Q. Cao, Z. Sun, dan H. Li, "Comparative Analysis of SARIMA, Prophet, and a Diagnostic Decomposition–Correction Hybrid for Long-Horizon Lottery Sales Forecasting," *Entropy*, vol. 28, no. 3, hlm. 286, Mar 2026, doi: 10.3390/e28030286.
- [14] H. N. Azalia, Hannie, dan A. A. Ridha, "ANALISIS TREN DAN FORECASTING KUNJUNGAN PASIEN BERBASIS FACEBOOK PROPHET UNTUK OPTIMALISASI MANAJEMEN LAYANAN KESEHATAN," *J. Inform. Dan Tek. Elektro Terap.*, vol. 14, no. 2, Apr 2026, doi: 10.23960/jitet.v14i2.9290.
- [15] F. B. Prakoso, G. Darmawan, dan A. Bachrudin, "Penerapan Metode Facebook Prophet Untuk Meramalkan Jumlah Penumpang Trans Metro Bandung KORIDOR 1," *ARMADA J. Penelit. Multidisiplin*, vol. 1, no. 3, hlm. 133–147, Mar 2023, doi: 10.55681/armada.v1i3.416.
- [16] M. R. Yudistira, "Nowcasting Jumlah Penumpang Udara di Kabupaten Berau Menggunakan Indeks Google Trends," *BESTARI BPS Kalimantan Timur*, vol. 5, no. 1, Jun 2025, Diakses: 20 Mei 2026. [Daring]. Tersedia pada: <https://bestari.bpskalim.com/index.php/bestari-bpskalim/article/view/68>
- [17] V. Ripatti, "Forecasting Prescription Medication Utilization: A Comparative Study of SARIMA, Prophet, XGBoost and LSTM Models".
- [18] M. G. S. Kenyi dan K. Yamamoto, "A hybrid SARIMA-Prophet model for predicting historical streamflow time-series of the Sobat River in South Sudan," *Discov. Appl. Sci.*, vol. 6, no. 9, hlm. 457, Agu 2024, doi: 10.1007/s42452-024-06083-x.
- [19] G. Saquare dan I. Beary, "Hybrid SARIMA-Facebook Prophet Model for Prediction and Forecasting of the Staple Food Prices in Tanzania," *Int. J. Oper. Res.*, vol. 20, hlm. 1–11, Agu 2023, doi: 10.6886/IJOR.202308_20.0001.
- [20] A. Lisnawati, S. Alam, dan I. Kurniawan, "Aplikasi Prediksi Kepadatan Wisatawan Mancanegara Ke Indonesia Dengan Algoritma Long Short-Term Memory (Lstm)," *Jati J. Mhs. Tek. Inform.*, vol. 9, no. 5, hlm. 8343–8349, Jul 2025, doi: 10.36040/jati.v9i5.15055.
- [21] D. J. Putnarubun, C. F. Palembang, J. E. T. Radjabaycolle, dan E. M. C. Wattimena, "Sistem Prediksi Waktu Keberangkatan Pesawat Berbasis Machine Learning," *J. Inform. Dan Tek. Elektro Terap.*, vol. 14, no. 2, Apr 2026, doi: 10.23960/jitet.v14i2.9320.
- [22] W. S. Kesuma, "Prediksi Analisis Time Series Nilai Tukar Rupiah terhadap Dolar Amerika dengan ARIMA dan SARIMA melalui CRISP-DM," *Pros. Semin. Nas. Univ. Ma Chung Inform. Sist. Inf. Bhs. Dan Seni Farm.*, vol. 4, hlm. 190–203, Sep 2024.
- [23] "KIP PT. Kereta Commuter Indonesia." Diakses: 16 Mei 2026. [Daring]. Tersedia pada: <https://kip.kci.id/info-berkala>
- [24] S. Pangestu, A. Febrian, dan Y. Cahyan, "PENERAPAN MODEL PROPHET UNTUK PERAMALAN JUMLAH DOKTER UMUM DI KOTA BANDUNG," *JATI J. Mhs. Tek. Inform.*, vol. 10, no. 1, hlm. 435–440, Jan 2026, doi: 10.36040/jati.v10i1.16619.
- [25] S. J. Taylor dan B. Letham, "Forecasting at scale," 27 September 2017, *PeerJ Preprints*. doi: 10.7287/peerj.preprints.3190v2.
- [26] A. A. Maulana dan H. Rosalina, "IMPLEMENTASI METODE SARIMAX UNTUK PREDIKSI CURAH HUJAN JANGKA PENDEK DI PAGERAGEUNG, TASIKMALAYA," *J. SUMBER DAYA AIR*, vol. 20, no. 1, hlm. 39–50, Mei 2024, doi: 10.32679/jsda.v20i1.874.
- [27] A. S. S. Ayuningrum, N. C. A. P. Sari, N. F. Ihsan, dan Nasrudin -, "Peramalan Jumlah Wisatawan Mancanegara yang Datang ke Bali Tahun 2025 Menggunakan SARIMAX dan Data Google Trends," *Semin. Nas. Off. Stat.*, vol. 2025, no. 1, hlm. 287–296, Des 2025, doi: 10.34123/semnasoffstat.v2025i1.2472.
- [28] M. G. S. Kenyi dan K. Yamamoto, "A hybrid SARIMA-Prophet model for predicting historical streamflow time-series of the Sobat River in South Sudan," *Discov. Appl. Sci.*, vol. 6, no. 9, hlm. 457, Agu 2024, doi: 10.1007/s42452-024-06083-x.
- [29] R. Achmad Zahra dan D. H. Fudholi, "Prediksi Pencapaian Target Kerja untuk Optimasi Manajemen Proyek Menggunakan Metode Long Short Term Memory dan Random Forest," *J. Sains Dan Teknol. J. Keilmuan Dan Apl. Teknol. Ind.*, vol. 25, no. 2, hlm. 208–222, 2025, doi: 10.36275/8m9pg216.
- [30] K. Arunika dan L. J. E. Dewi, "Perbandingan Model Sarima, Exponential Smoothing, Dan Xgboost Untuk Prediksi Penjualan Super Store," *J. Inform. Dan Tek. Elektro Terap.*, vol. 13, no. 3S1, Okt 2025, doi: 10.23960/jitet.v13i3S1.8173.