

Peramalan Jumlah Penumpang Kereta Api Menggunakan Metode Arima dan *Grey System Theory*

Ria Susilawati, Siti Sunendiari*

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

ARTICLE INFO

Article history :

Received : 2/4/2022
Revised : 6/7/2022
Published : 7/7/2022



Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Volume : 2
No. : 1
Halaman : 1-13
Terbitan : Juli 2022

ABSTRAK

Metode peramalan kuantitatif dibagi menjadi tiga jenis, yaitu metode deret waktu, metode kausal, dan metode kombinasi metode deret waktu dan metode kausal (Makridakis, 1999). Metode deret waktu adalah sebuah metode yang memperkirakan masa depan berdasarkan nilai masa lalu dari suatu variabel atau kesalahan masa lalu. Salah satu metode deret waktu yang sedang berkembang dan umum digunakan adalah metode ARIMA. Meskipun proses perhitungannya cukup kompleks, metode ARIMA akan memberikan hasil yang lebih baik dari metode-metode peramalan lainnya, karena metode ini tidak mengabaikan kaidah-kaidah pada deret waktu seperti pengujian stasioner, penaksiran parameter, dan pemeriksaan diagnostik. Berbeda dengan metode ARIMA, *Grey System Theory* tidak memerlukan seluruh data historis, namun hanya membutuhkan minimal empat data historis dan dalam interval yang sama. Dalam penelitian ini akan dibahas Peramalan Jumlah Penumpang Kereta Api Wilayah Jawa (Jabodetabek dan Non Jabodetabek) menggunakan Metode ARIMA dan *Grey System Theory*. Dari pemodelan yang telah dilakukan diperoleh hasil bahwa model GM (1,1) menjadi model yang paling cocok digunakan untuk meramalkan jumlah penumpang kereta api wilayah Jawa.

Kata Kunci : Metode Peramalan; ARIMA; Grey System Theory

ABSTRACT

Quantitative forecasting methods are divided into three types, its time series methods, causal methods, and combination methods of time series and causal methods (Makridakis, 1999). The time series method is a method that predicts the future based on the past value of a variable or past errors. One of the time series methods that is being developed and commonly used is the ARIMA method. Although the calculation process is quite complex, the ARIMA method will give better results than other forecasting methods, because this method does not ignore time series rules such as stationary testing, parameter estimation, and diagnostic testing. Unlike the ARIMA method, *Grey System Theory* does not require all historical data, but only requires a minimum of four historical data and in the same interval. In this study, we will discuss the Forecasting of the Number of Railway Passengers in the Java Region (Jabodetabek and Non Jabodetabek) using the ARIMA Method and *Grey System Theory*. From the modeling that has been done, it is found that the GM model (1,1) is the most suitable model to be used to predict the number of train passengers in the Java region.

Keywords : Forecasting Methods; ARIMA; Grey System Theory

© 2022 Jurnal Riset Statistika Unisba Press. All rights reserved.

A. Pendahuluan

Pada dasarnya terdapat dua metode peramalan, yaitu metode peramalan kuantitatif dan metode peramalan kualitatif. Metode kuantitatif merupakan metode peramalan yang digunakan ketika data historis tersedia [1]. Metode kuantitatif dibagi menjadi tiga jenis, yaitu metode deret waktu, metode kausal, dan kombinasi metode deret waktu dan metode kausal. Sedangkan metode kualitatif adalah metode peramalan yang digunakan ketika data historis tidak tersedia dan bersifat subjektif atau intuitif. Metode ini lebih menekankan pada keputusan-keputusan hasil diskusi pendapat pribadi seseorang dan intuisi meskipun kelihatannya kurang ilmiah tetapi dapat memberikan hasil yang baik [3].

Metode deret waktu adalah suatu metode kuantitatif untuk menentukan pola data masa lampau yang telah dikumpulkan secara teratur [4]. Metode ARIMA merupakan salah satu metode deret waktu yang sedang berkembang dan umum digunakan yang diperkenalkan dan dikembangkan oleh George Box dan Gwilym Jenkins pada tahun 1960-an. Metode ARIMA adalah metode yang tidak mengabaikan kaidah-kaidah deret waktu seperti pengujian stasioneritas, penaksiran parameter, dan pemeriksaan diagnostik. Jika dibandingkan dengan metode-metode peramalan lainnya proses perhitungan metode ARIMA terbilang cukup kompleks. Meskipun begitu metode ARIMA (*Box-Jenkins*) dapat memberikan hasil yang lebih baik.

Grey System Theory ditemukan oleh Prof. Julong Deng pada tahun 1982. *Grey System Theory* ini tidak memerlukan seluruh data historis, namun hanya membutuhkan minimal empat data historis dan dalam interval yang sama dan metode ini berfokus pada peramalan dengan sampel data kecil [5]. Sedangkan untuk membangun model ARIMA, Box dan Jenkins menyarankan ukuran sampel yang dibutuhkan adalah 50 data pengamatan, terlebih lagi untuk data deret waktu yang bersifat musiman diperlukan ukuran sampel yang lebih besar lagi [6].

Pada penelitian sebelumnya, untuk memprediksi emisi CO₂ di Iran, penelitian tersebut telah menggunakan *Grey System Theory* dan ARIMA dengan cara membandingkan dua metode tersebut [7]. Penelitian tersebut melakukan peramalan selama 10 tahun kedepan dan memperoleh hasil bahwa metode *Grey System Theory* menghasilkan nilai MAPE, MAE, RMSE yang lebih kecil dibandingkan metode ARIMA [8]. Dalam skripsi Dian Kurnianingsih Pratiwi, metode ARIMA dan *Grey System Theory* digunakan untuk memprediksi jumlah wisatawan mancanegara ke Indonesia dengan cara membandingkan kedua model yang diperoleh dari dua metode tersebut.

Berdasarkan uraian di atas untuk meramalkan jumlah penumpang kereta api wilayah Jawa (Jabodetabek dan non Jabodetabek) dalam penelitian ini akan diterapkan metode ARIMA dan *Grey System Theory*. Data yang digunakan adalah data bulanan jumlah penumpang kereta api wilayah Jawa dari bulan Januari 2016 sampai dengan bulan Desember 2020. Untuk mengetahui model yang paling cocok digunakan untuk meramalkan jumlah penumpang kereta api wilayah Jawa, maka akan dilakukan perbandingan nilai MAPE dari kedua model metode peramalan tersebut.

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini adalah: (1) Bagaimana model ARIMA dalam meramalkan jumlah penumpang kereta api wilayah Jawa pada bulan Januari 2016 – Desember 2020?; (2) Bagaimana model *Grey System Theory* dalam meramalkan jumlah penumpang kereta api wilayah Jawa pada bulan Januari 2016 – Desember 2020?; (3) Diantara model ARIMA dan *Grey System Theory* manakah model yang paling cocok untuk meramalkan jumlah penumpang kereta api wilayah Jawa pada bulan Januari 2021 – Desember 2021?

Selanjutnya, tujuan dari penelitian ini diuraikan sebagai berikut: (1) Untuk mengetahui model-ARIMA dalam meramalkan jumlah penumpang kereta api wilayah Jawa ada bulan Januari 2016 – Desember 2020; (2) Untuk mengetahui model *Grey System Theory* dalam meramalkan jumlah penumpang kereta api wilayah Jawa pada bulan Januari 2016 – Desember 2020; (3) Untuk mengetahui model yang paling cocok diantara model ARIMA dan *Grey System Theory* untuk meramalkan jumlah penumpang kereta api wilayah Jawa pada bulan Januari 2021 – Desember 2021.

B. Metode Penelitian

Data diperoleh dari Badan Pusat Statistik Indonesia melalui akses internet pada <https://www.bps.go.id/>. Data dibagi menjadi 2 bagian yaitu sebesar 80% data digunakan sebagai data *training* dan 20% data digunakan

sebagai data *testing* dimana jumlah pengamatan adalah 60 bulan [9]. Data *training* yaitu data selama periode Januari 2016 sampai Desember 2019 digunakan untuk proses pemodelan, sementara data *testing* yaitu data selama periode Januari 2020 sampai Desember 2020 digunakan untuk proses perhitungan MAPE.

Prosedur Penelitian

Pertama, pemodelan dilakukan dengan menggunakan metode ARIMA. Berikut ini langkah-langkah pemodelan menggunakan ARIMA: (1) Melakukan identifikasi model untuk melihat kestasioneran data dengan cara memplot data *training*. Atau dengan cara menggunakan uji akar unit *Augmented Dickey-Fuller Test* sebagai berikut:

$$\tau = \frac{\hat{\delta}}{se(\hat{\delta})} \tag{1}$$

Kedua, Jika data sudah stasioner, untuk memperkirakan orde AR dan MA yang akan diambil maka hitung dan telaah ACF dan PACF data *training*, namun jika data belum stasioner, untuk memperoleh data stasioner maka lakukan proses *differencing* terhadap data *training* dengan menggunakan:

$$(1-B)^d Y_t \tag{2}$$

Ketiga, Apabila data *training* sudah stasioner, untuk memperkirakan orde AR dan MA yang akan diambil dengan menggunakan data *training* kemudian plot ACF dan PACF. Keempat, Berdasarkan orde AR dan MA yang akan diambil, maka tetapkan model tentatif ARIMA yang mungkin terbentuk. Kelima, Melakukan pengujian keberartian dari parameter AR(*p*) dan MA(*q*) untuk setiap model ARIMA yang mungkin terbentuk dengan menggunakan :

$$t = \frac{\hat{\phi}}{se(\hat{\phi})} \tag{3}$$

$$t = \frac{\hat{\theta}}{se(\hat{\theta})} \tag{4}$$

Keenam, Melakukan pengujian diagnostik dengan menggunakan:

$$Q = n(n + 2) \sum_{k=1}^K (n - k)^{-1} \hat{\rho}_k^2 \tag{5}$$

Ketujuh, Melakukan peramalan untuk data *testing*. Kedelapan, melakukan perhitungan MAPE dengan menggunakan data *testing* dan menggunakan rumus :

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - F_t}{X_t} \right| \times 100 \tag{6}$$

Dan terakhir melakukan pemilihan model terbaik berdasarkan hasil perhitungan MAPE. Selanjutnya, pemodelan menggunakan metode *Grey System Theory* untuk menentukan GM (1,1). Berikut ini langkah-langkah menentukan GM (1,1):

Pertama, Mendefinisikan barisan data asli, yaitu data *training* berdasarkan urutan waktunya. Dengan menggunakan :

$$Y^{(0)} = (y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n)), \quad n \geq 4 \tag{7}$$

Kedua, melakukan perhitungan *first accumulating generation operation* (1-AGO) atau pembangkit operasi akumulasi. Dengan menggunakan :

$$Y^{(1)}(k) = \sum_{i=1}^k y^{(0)}(i), \quad k = 1, 2, 3, \dots, n \tag{8}$$

Ketiga, melakukan perhitungan nilai rata-rata dari dua data $Y^{(1)}$ yang berdekatan.

$$Z^{(1)}(k) = \frac{1}{2}y^{(1)}(k) + \frac{1}{2}y^{(1)}(k - 1), \quad k = 2, 3, \dots, n \tag{9}$$

Keempat, melakukan perhitung nilai parameter a dan b pada GM (1,1). Dengan a dinotasikan sebagai developing coefficient dan b adalah grey input (Deng, 1989). Berikut persamaan yang digunakan untuk menghitung nilai parameter a dan b:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y \tag{10}$$

Kelima, menghitung respon waktu dari GM (1,1) setelah mengetahui nilai parameter a dan b dengan menggunakan persamaan berikut:

$$y_p^{(1)}(k + 1) = \left[y^{(0)}(1) - \frac{b}{a} \right] e^{-ak} + \frac{b}{a} \tag{11}$$

Keenam, $y_p^{(0)}(k + 1)$ merupakan notasi dari nilai peramalan dalam metode GM (1,1). Inverse Accumulated Generating Operation (IAGO) digunakan untuk menghitung nilai peramalan dengan menggunakan persamaan berikut:

$$y_p^{(0)}(k + 1) = \left[y^{(0)}(1) - \frac{b}{a} \right] e^{-ak} (1 - e^a) \tag{12}$$

Terakhir, melakukan perhitungan MAPE dengan menggunakan data *testing* dan menggunakan rumus :

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - F_t}{X_t} \right| \times 100 \tag{13}$$

Setelah mendapatkan nilai MAPE dari kedua model, yaitu model ARIMA dan model *Grey System Theory* bandingkan nilai MAPE dari kedua model tersebut. Setelah memperoleh model terbaik, lakukan peramalan untuk periode selanjutnya menggunakan model terbaik yang diperoleh tersebut.

C. Hasil dan Pembahasan

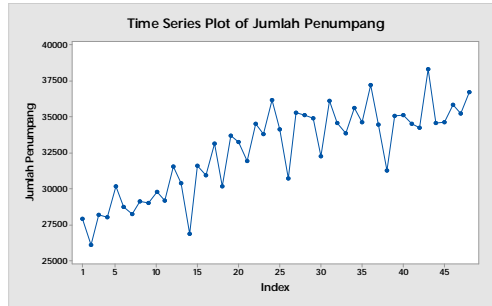
Hasil Pemodelan dan Peramalan Menggunakan Metode ARIMA

Langkah pertama adalah membuat plot data training untuk melihat apakah data sudah stasioner atau belum. Data disajikan pada Tabel 1.

Tabel 1. Data Training

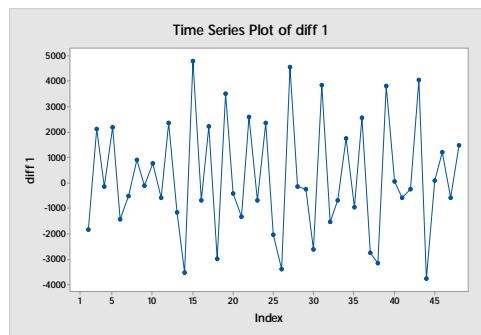
No	Bulan	Jumlah Penumpang Kereta Api Wilayah Jawa
1	Januari 2016	27886
2	Februari 2016	26058
3	Maret 2016	28156
4	April 2016	28000
5	Mei 2016	30176
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
48	Desember 2019	36710

Berikut diperoleh grafik deret waktu data *training* :



Gambar 1. Grafik Data *Training*

Berdasarkan Gambar 1 terlihat bahwa adanya trend naik terhadap data *training* sehingga hal tersebut menunjukkan bahwa data *training* tidak stasioner dalam rata-rata. Maka data tersebut harus distasionerkan melalui proses stasioneritas, yaitu *differencing* atau pembedaan. Proses *differencing* dilakukan agar mendapatkan data yang stasioner sehingga dapat digunakan ke tahap selanjutnya. Diperoleh grafik hasil *differencing* pertama sebagai berikut.



Gambar 2. Grafik Data *Training* Hasil *Differencing* Pertama

Berdasarkan Gambar 2 menunjukkan bahwa pola data berada pada kesetimbangan di sekitar nilai rata-rata yang konstan dan varians di sekitar rata-rata tersebut konstan selama waktu tertentu, ini menunjukkan bahwa data tersebut sudah stasioner. Untuk lebih memastikan bahwa data tersebut sudah stasioner, maka dilakukan pengujian secara formal melalui uji akar unit (*Augmented Dickey-Fuller-Test*) menggunakan *software* Eviews versi 10, mengingat metode visual cenderung bersifat subjektif.

Hasil uji akar unit data *training* hasil *differencing* pertama menggunakan metode *Augmented Dickey-Fuller* disajikan pada Tabel 2 dan Tabel 3.

Tabel 2. Uji ADF Data *Training* Hasil *Differencing* Pertama Tanpa Intersep

	t-Statistic	Prob.
<i>Augmented Dickey-Fuller</i>	-7.040791	0.0000

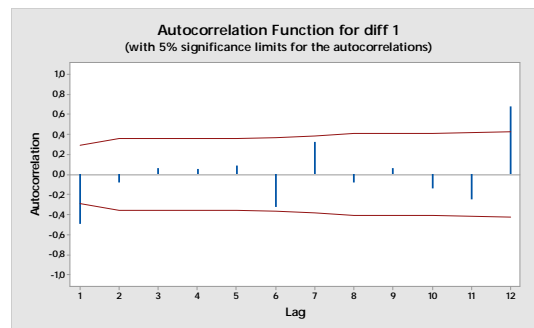
Diperoleh nilai *t-statistic* sebesar -7.040791 dan nilai probabilitas *ADF-Test* sebesar 0.0000. Dengan taraf signifikan (α) sebesar 5% maka dapat disimpulkan bahwa nilai probabilitas *ADF-Test* lebih kecil dari taraf signifikan (α) 5%, maka H_0 ditolak yang berarti tidak terdapat unit roots atau data *training* hasil *differencing* pertama stasioner dalam rata-rata tanpa intersep.

Tabel 3. Uji ADF Data *Training* Hasil *Differencing* Pertama Dengan Intersep

	t-Statistic	Prob.
<i>Augmented Dickey-Fuller</i>	-7.598400	0.0000
C	2.060064	0.0459

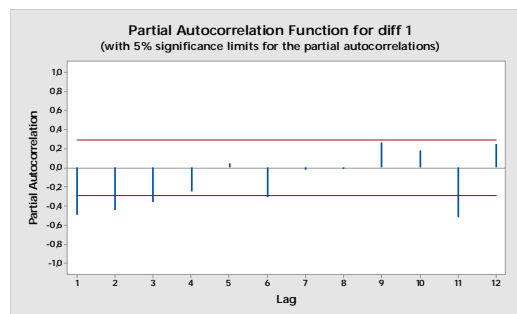
Pada variabel C (intersep) diperoleh nilai probabilitas sebesar 0.0459 yang berarti lebih kecil dari taraf signifikan sebesar 5%, sehingga intersep mempengaruhi ADF. Kemudian diperoleh nilai t-Statistic sebesar -7.598400 dan nilai probabilitas ADF-Test sebesar 0.0000 yang berarti nilai probabilitas ADF-Test lebih kecil dari taraf signifikan sebesar 5%, maka H_0 ditolak yang berarti tidak terdapat unit roots atau data training hasil differencing pertama stasioner dalam rata-rata dengan intersep.

Plot ACF dan PACF digunakan untuk memperkirakan orde autoregresi dan rata-rata bergerak yang akan diambil yang kemudian digunakan untuk membuat model tentatif ARIMA. Nilai ACF dan PACF hingga lag ke-12 disajikan dalam Gambar 3 dan Gambar 4.



Gambar 3. *Correlogram* ACF Data *Training* Hasil *Differencing* Pertama

Berdasarkan *correlogram* ACF data *training* hasil *differencing* pertama terlihat bahwa nilai ACF signifikan pada lag ke-1 dan cut off setelah lag ke-1. Sehingga diperoleh MA (1) dan MA(2).



Gambar 4. *Correlogram* PACF Data *Training* Hasil *Differencing* Pertama

Sedangkan untuk *correlogram* PACF data *training* hasil *differencing* pertama terlihat bahwa nilai PACF signifikan pada lag ke-1 kemudian turun secara perlahan sampai lag ke-4 dan cut-off setelah lag ke-4. Sehingga diperoleh AR(1), AR(2), AR(3) dan AR (4). Maka model tentatif ARIMA yang terbentuk adalah sebagai berikut : (1) Model ARIMA (1,1,1); (2) Model ARIMA (0,1,1); (3) Model ARIMA (1,1,0); (4) Model ARIMA (0,1,2); (5) Model ARIMA (1,1,2); (6) Model ARIMA (2,1,0); (7) Model ARIMA (2,1,1); (8) Model ARIMA (2,1,2); (9) Model ARIMA (3,1,0); (10) Model ARIMA (3,1,1); (11) Model ARIMA (3,1,2); (12) Model ARIMA (4,1,0); (13) Model ARIMA (4,1,1); (14) Model ARIMA (4,1,2).

Penaksiran Parameter Model

Diperoleh hasil bahwa model ARIMA (0,1,1), ARIMA (1,1,0), ARIMA (1,1,2) dan ARIMA (2,1,0) yang memungkinkan layak untuk digunakan yang disajikan dalam tabel 4 di bawah ini

Tabel 4. Pengujian Parameter Model ARIMA

Model	Parameter	T-hitung	T-tabel	P-Value	Keputusan
ARIMA(0,1,1)	MA(1)	-5,587873	$t_{(0,025,47)}=2,011741$	0,0000	Signifikan
ARIMA(1,1,0)	AR(1)	-3,913666	$t_{(0,025,47)}=2,011741$	0,0000	Signifikan
ARIMA(1,1,2)	AR(1)	5,191644	$t_{(0,025,47)}=2,011741$	0,0000	Signifikan
	MA(1)	-14,13430		0,0000	Signifikan
	MA(2)	8,370511		0,0000	Signifikan
ARIMA(2,1,0)	AR(1)	-3,918358	$t_{(0,025,47)}=2,011741$	0,0003	Signifikan
	AR(2)	-3,007100		0,0043	Signifikan

Pemeriksaan Diagnostik

Pemeriksaan Diagnostik Model ARIMA (0,1,1). Dengan nilai $n = 48$, maka Q adalah:

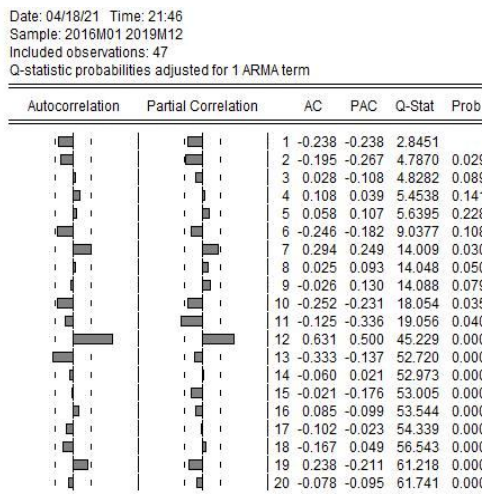
$$Q = 48(48 + 2) \left(\frac{-0,238^2 + \dots + -0,078^2}{48 - 20} \right) \tag{14}$$

$$Q = 82,2171$$

Diperoleh nilai $\chi^2_{(\alpha;K-p-q)} = \chi^2_{(0,05;20-0-1)} = 30,1$

Jadi dapat disimpulkan bahwa nilai $Q > \chi^2_{tabel}$ sehingga H_0 ditolak. Artinya ada korelasi antar residual, rangkaian data residual tidak acak, dan ada pola dalam deret residual ini menandakan bahwa residual tidak memenuhi syarat white noise sehingga bisa dikatakan bahwa model ARIMA (0,1,1) tidak layak digunakan untuk peramalan.

Pada Gambar 3 terlihat bahwa terdapat nilai probabilitas lag model ARIMA (0,1,1) yang nilainya lebih kecil dari $\alpha = 0,05$ artinya terdapat nilai probabilitas yang signifikan, sehingga H_0 ditolak yang artinya model tidak memenuhi syarat cukup (residual tidak memenuhi syarat white noise atau rangkaian data tidak acak). Maka dari itu, dapat dikatakan bahwa model ARIMA (0,1,1) tidak layak digunakan untuk peramalan.



Gambar 3. Correlogram ACF dan PACF Residual Model ARIMA (0,1,1)

Pemeriksaan Diagnostik Model ARIMA (1,1,0). Dengan nilai $n = 48$, maka Q adalah:

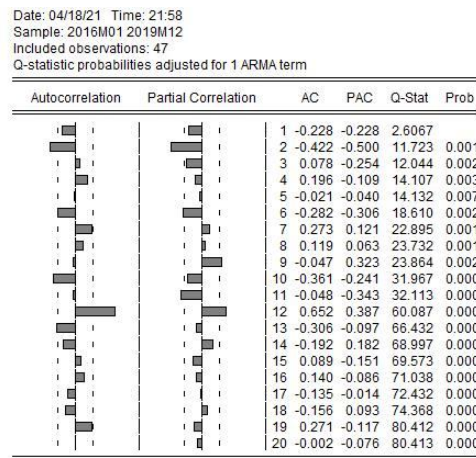
$$Q = 48(48 + 2) \left(\frac{-0,228^2 + \dots + -0,002^2}{48-20} \right) \tag{15}$$

$$Q = 109,4732571$$

Diperoleh nilai $\chi^2_{(\alpha;K-p-q)} = \chi^2_{(0,05;20-0-1)} = 30,1$

Jadi dapat disimpulkan bahwa nilai $Q > \chi^2_{tabel}$ sehingga H_0 ditolak. Artinya ada korelasi antar residual, rangkaian data residual tidak acak, dan ada pola dalam deret residual ini menandakan bahwa residual tidak memenuhi syarat *white nose* sehingga bisa dikatakan bahwa model ARIMA (1,1,0) tidak layak digunakan untuk peramalan.

Pada Gambar 4 terlihat bahwa nilai probabilitas disemua lag model ARIMA (1,1,0) lebih kecil dari $\alpha = 0,05$ artinya ada nilai probabilitas yang signifikan, sehingga H_0 ditolak yang artinya model tidak memenuhi syarat cukup (residual tidak memenuhi syarat *white nose* atau rangkaian data tidak acak). Maka dari itu, dapat dikatakan bahwa model ARIMA (1,1,0) tidak layak digunakan untuk peramalan.



Gambar 4. Correlogram ACF dan PACF Residual Model ARIMA (1,1,0)

Pemeriksaan Diagnostik Model ARIMA (1,1,2) Dengan nilai $n = 48$, maka Q adalah:

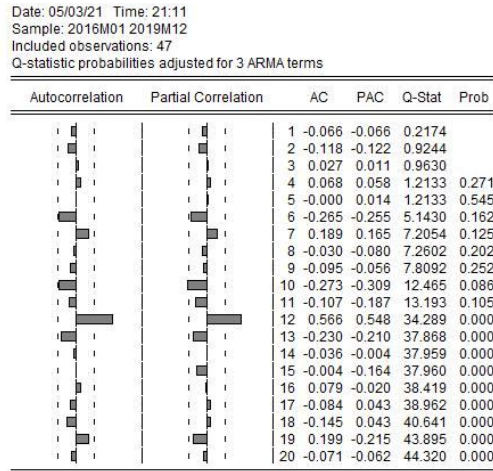
$$Q = 48(48 + 2) \left(\frac{-0,066^2 + \dots + -0,071^2}{48-20} \right) \tag{16}$$

$$Q = 58,2012$$

Diperoleh nilai $\chi^2_{(\alpha;K-p-q)} = \chi^2_{(0,05;20-0-1)} = 30,1$

Jadi dapat disimpulkan bahwa nilai $Q > \chi^2_{tabel}$ sehingga H_0 ditolak. Artinya ada korelasi antar residual, rangkaian data residual tidak acak, dan ada pola dalam deret residual ini menandakan bahwa residual tidak memenuhi syarat *white nose* sehingga bisa dikatakan bahwa model ARIMA (1,1,2) tidak layak digunakan untuk peramalan.

Pada Gambar 5 terlihat bahwa terdapat nilai probabilitas lag model ARIMA (1,1,2) yang nilainya lebih kecil dari $\alpha = 0,05$ artinya terdapat nilai probabilitas yang signifikan, sehingga H_0 ditolak yang artinya model tidak memenuhi syarat cukup (residual tidak memenuhi syarat *white nose* atau rangkaian data tidak acak). Maka dari itu, dapat dikatakan bahwa model ARIMA (1,1,2) tidak layak digunakan untuk peramalan.



Gambar 5. Correlogram ACF dan PACF Residual Model ARIMA (1,1,2)

Pemeriksaan diagnostik Model ARIMA (2,1,0). Dengan nilai $n = 48$, maka Q adalah:

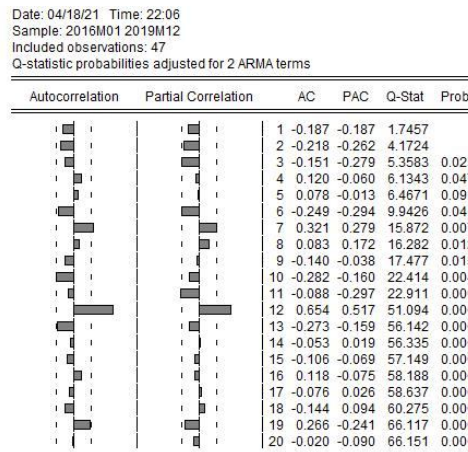
$$Q = 48(48 + 2) \left(\frac{-0,187^2 + \dots + -0,020^2}{48 - 20} \right) \tag{17}$$

$$Q = 86,80962857$$

Diperoleh nilai $\chi^2_{(\alpha;K-p-q)} = \chi^2_{(0,05;20-0-1)} = 30,1$

Jadi dapat disimpulkan bahwa nilai $Q > \chi^2_{tabel}$ sehingga H_0 ditolak. Artinya ada korelasi antar residual, rangkaian data residual tidak acak, dan ada pola dalam deret residual ini menandakan bahwa residual tidak memenuhi syarat *white noise* sehingga bisa dikatakan bahwa model ARIMA (2,1,0) tidak layak digunakan untuk peramalan.

Pada Gambar terlihat bahwa nilai probabilitas disemua lag model ARIMA (2,1,0) lebih kecil dari $\alpha = 0,05$ artinya ada nilai probabilitas yang signifikan, sehingga H_0 ditolak yang artinya model tidak memenuhi syarat cukup (residual tidak memenuhi syarat *white noise* atau rangkaian data tidak acak). Maka dari itu, dapat dikatakan bahwa model ARIMA (2,1,0) tidak layak digunakan untuk peramalan.



Gambar 6. Correlogram ACF dan PACF Residual Model ARIMA (2,1,0)

Dikarenakan tidak ada model yang layak digunakan untuk peramalan maka tidak dapat dilanjutkan ke langkah berikutnya.

Hasil Pemodelan dan Peramalan menggunakan Metode *Grey System Theory*

Langkah pertama yang dilakukan dalam pemodelan menggunakan GM (1,1) adalah mendefinisikan barisan data asli berdasarkan urutan waktunya yang dinotasikan sebagai $Y^{(0)}$ menggunakan Persamaan (2.51). Dalam hal ini data yang digunakan adalah data *training*. Diperoleh hasil $Y^{(0)}$ sebagai berikut:

$$Y^{(0)} = (27.886; 26.058; 28.156; \dots; 36.710) \tag{18}$$

Dimana

$$y^{(0)}(1) = 27.886$$

$$y^{(0)}(2) = 26.058$$

⋮

$$y^{(0)}(48) = 36.710$$

Langkah kedua adalah melakukan perhitungan pembangkit operasi kumulasi atau first accumulating generation operation (1-AGO) yang dinotasikan sebagai $Y^{(1)}$.

Perhitungan nilai $y^{(1)}(1)$

$$y^{(1)}(1) = \sum_{i=1}^1 y^{(0)}(i) = y^{(0)}(1) = 27.886$$

Perhitungan nilai $y^{(1)}(2)$

$$y^{(1)}(2) = \sum_{i=1}^2 y^{(0)}(i)$$

$$y^{(1)}(2) = y^{(0)}(1) + y^{(0)}(2)$$

$$y^{(1)}(2) = 27.886 + 26.058$$

$$y^{(1)}(2) = 53.944$$

Begitu seterusnya dengan menggunakan rumus dan cara yang sama, perhitungan dilanjutkan sampai diperoleh hasil untuk $y^{(1)}(48)$. Langkah selanjutnya adalah melakukan perhitungan nilai rata-rata dari dua data $Y^{(1)}$ yang berdekatan. Hasil perhitungan ini dinotasikan sebagai $Z^{(1)}$.

Perhitungan nilai $z^{(1)}(2)$

$$z^{(1)}(2) = \frac{1}{2}y^{(1)}(2) + \frac{1}{2}y^{(1)}(2 - 1)$$

$$z^{(1)}(2) = \frac{1}{2}y^{(1)}(2) + \frac{1}{2}y^{(1)}(1)$$

$$z^{(1)}(2) = \frac{1}{2}(53.944) + \frac{1}{2}(27.886)$$

$$z^{(1)}(2) = 40.915$$

Perhitungan nilai $z^{(1)}(3)$

$$z^{(1)}(3) = \frac{1}{2}y^{(1)}(3) + \frac{1}{2}y^{(1)}(3 - 1)$$

$$z^{(1)}(3) = \frac{1}{2}y^{(1)}(3) + \frac{1}{2}y^{(1)}(2)$$

$$z^{(1)}(3) = \frac{1}{2}(82.100) + \frac{1}{2}(53.944)$$

$$z^{(1)}(3) = 68.022$$

Begitu seterusnya dengan menggunakan rumus dan cara yang sama, perhitungan dilanjutkan sampai diperoleh hasil untuk $z^{(1)}(48)$. Langkah keempat, yaitu menghitung nilai parameter a dan b dihitung dengan menggunakan Persamaan sebagai berikut:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y$$

Dengan

$$Y = \begin{bmatrix} 26.058 \\ 28.156 \\ \vdots \\ 36.170 \end{bmatrix}$$

$$B = \begin{bmatrix} -40.915 & 1 \\ -68.022 & 1 \\ \vdots & \vdots \\ -1.548.068 & 1 \end{bmatrix}$$

Jadi, diperoleh nilai parameter a sebesar 0,361049157 dan nilai parameter b sebesar -0,648579902. Selanjutnya dilakukan perhitungan respon waktu GM (1,1) sebagai berikut:

Perhitungan $y_p^{(1)}(2)$

$$y_p^{(1)}(2) = \left[27.886 - \frac{28580,81908}{-0,00543934} \right] e^{-(-0,00543934)1} + \frac{28580,81908}{-0,00543934}$$

$$y_p^{(1)}(2) = 56.696,78531$$

Perhitungan $y_p^{(1)}(3)$

$$y_p^{(1)}(3) = \left[27.886 - \frac{28580,81908}{-0,00543934} \right] e^{-(-0,00543934)2} + \frac{28580,81908}{-0,00543934}$$

$$y_p^{(1)}(3) = 86.664,70925$$

Begitu seterusnya dengan menggunakan rumus dan cara yang sama, perhitungan dilanjutkan sampai dengan diperoleh hasil untuk $y_p^{(1)}(49)$. Langkah selanjutnya adalah melakukan peramalan untuk data testing. Hasil peramalan ini selanjutnya digunakan untuk menghitung nilai MAPE untuk model GM (1,1). Nilai peramalan dihitung dengan Inverse Accumulated Generating Operation (IAGO) sebagai berikut:

$$y_p^{(0)}(k + 1) = \left[y^{(0)}(1) - \frac{28580,81908}{-0,00543934} \right] e^{-(-0,00543934)k} (1 - e^{-0,00543934})$$

Hasil peramalan untuk data testing disajikan pada Tabel 5 sebagai berikut:

Tabel 5. Hasil Peramalan Untuk Data *Testing* Menggunakan Model GM (1,1)

Model	Bulan	Periode	Y_t	Hasil Peramalan
GM (1,1)	Januari 2020	49	33472	37.203
	Februari 2020	50	31679	37.406
	Maret 2020	51	22949	37.610
	April 2020	52	5813	37.815
	Mei 2020	53	5476	38.022
	Juni 2020	54	9272	38.229
	Juli 2020	55	12205	38.438
	Agustus 2020	56	12679	38.647
	September 2020	57	11295	38.858
	Oktober 2020	58	11768	39.070

Lanjutan Tabel 5. Hasil Peramalan Untuk Data *Testing* Menggunakan Model GM (1,1)

Model	Bulan	Periode	Y_t	Hasil Peramalan
GM (1,1)	Nopember 2020	59	13523	39.283
	Desember 2020	60	13262	39.497

Dikarenakan tidak ada model ARIMA yang layak digunakan untuk peramalan maka tidak ada perhitungan MAPE. Nilai MAPE digunakan sebagai nilai pembandingan untuk kedua model, yaitu model ARIMA dan model GM (1,1). Dengan tidak adanya model ARIMA yang layak digunakan untuk peramalan maka dapat disimpulkan bahwa model GM (1,1) menjadi model yang paling cocok digunakan untuk meramalkan jumlah penumpang kereta api wilayah Jawa (Jabodetabek dan non Jabodetabek).

Peramalan Jumlah Penumpang Kereta Api Wilayah Jawa (Jabodetabek dan Non Jabodetabek)

Dilakukan peramalan jumlah penumpang kereta api wilayah Jawa untuk bulan Januari 2021 – Desember 2021 menggunakan model GM (1,1) sebagai berikut:

Tabel 6. Hasil Peramalan Jumlah Penumpang Kereta Api Wilayah Jawa

Bulan	Hasil Peramalan
Januari 2021	39.713
Februari 2021	39.929
Maret 2021	40.147
April 2021	40.366
Mei 2021	40.586
Juni 2021	40.808
Juli 2021	41.030
Agustus 2021	41.254
September 2021	41.479
Oktober 2021	41.705
November 2021	41.933
Desember 2021	42.161
Jumlah	491.111

Terlihat bahwa hasil peramalan jumlah penumpang kereta api wilayah Jawa untuk bulan Januari 2021 – Desember 2021 setiap bulannya mengalami kenaikan. Berdasarkan hasil peramalan jumlah penumpang kereta api wilayah Jawa di tahun 2021 mencapai 491.111 penumpang.

D. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, maka dapat ditarik beberapa kesimpulan yang dapat dikemukakan seperti, dari pemodelan yang telah dilakukan menggunakan ARIMA untuk data jumlah penumpang kereta api wilayah Jawa (Jabodetabek dan non Jabodetabek) pada bulan Januari 2016 – Desember 2020 diperoleh empat belas model ARIMA yang mungkin terbentuk. Setelah dilakukan penaksiran parameter model dan pemeriksaan diagnostik untuk keempat belas model ARIMA yang mungkin terbentuk diperoleh hasil bahwa tidak ada model ARIMA yang cocok untuk digunakan dalam peramalan.

Dari pemodelan *Grey System Theory* yang telah dilakukan untuk data jumlah penumpang kereta api wilayah Jawa pada bulan Januari 2016 – Desember 2020 menggunakan *Grey Model* (1,1) atau GM (1,1) diperoleh nilai parameter a sebesar -0,00543934 dan nilai parameter b sebesar 28580, 81908. Sehingga diperoleh persamaan untuk melakukan peramalan menggunakan *Inverse Accumulated Generating Operation* (IAGO) sebagai berikut:

$$y_p^{(0)}(k + 1) = \left[y^{(0)}(1) - \frac{28580,81908}{-0,00543934} \right] e^{-(-0,00543934)k} (1 - e^{-0,00543934})$$

Karena tidak ada model ARIMA yang cocok digunakan untuk peramalan maka tidak ada perbandingan nilai MAPE sebagai pembandingan kedua model, yaitu model ARIMA dan model GM (1,1). Oleh karena itu, model GM (1,1) adalah model yang paling cocok digunakan untuk meramalkan jumlah penumpang kereta api wilayah Jawa.

Daftar Pustaka

- [1] B. Nugroho, *Pengantar Statistik*, 4th ed. Yogyakarta: UPP AMP YKPN, 2001.
- [2] Sheila Ghazia Shofwani and A. Kudus, "Penentuan Kriteria Pengunjung dalam Pemilihan Green Hotel di Kota Bandung Menggunakan Metode Discrete Choice Experiment dengan Desain Choice Sets Kombinatorial," *J. Ris. Stat.*, vol. 1, no. 1, pp. 1–9, Jul. 2021, doi: 10.29313/jrs.v1i1.5.
- [3] M. S. G, S. W. C, and V. M. E, *Metode dan Aplikasi Peramalan*, 1st ed. Jakarta: Erlangga, 1991.
- [4] B. L. Bowerman and R. T. O'Connell, *Time Series Forecasting*. Boston: Duxbury Press, 1987.
- [5] D. Julong, "Introduction to grey systems theory," *J. Grey Syst.*, pp. 1–24, 1988, doi: 10.1007/978-3-642-16158-2_1.
- [6] Y. T. S, *Analisis Deret Waktu*. Bandung: Pustaka Ceria, 2010.
- [7] M. R. Lotfalipour, M. A. Falahi, and M. Bastam, "Prediction of CO2 emissions in Iran using grey and ARIMA models," *Int. J. Energy Econ. Policy*, vol. 3, no. 3, pp. 229–237, 2013.
- [8] G. F. Fitri, F. Agustina, and R. Marwati, "Penerapan Metode Grey System pada Peramalan Produk Olifens (Studi Kasus PT. Chandra Asri Petrochemical Tbk)," *EurekaMatika*, vol. 6, pp. 52–63, 2018.
- [9] G. Felici and L. Tosheff, *Mathematical Methods for Knowledge Discovery and Data Mining*. *Information science reference Library of Congress Cataloging-in-Publication Data*. 2008. [Online]. Available: <http://www.igi-global.com/reference%5Cnhttp://www.eurospanonline.com>