

IMPLEMENTASI *TIME SERIES* PADA DATA PENJUALAN DI GAIKINDO MENGGUNAKAN ALGORITMA *SEASONAL ARIMA*

Muhammad Heru Widiyanto, Rini Mayasari, Garno

Program Studi Informatika S1, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Indonesia
1910631170213@student.unsika.ac.id

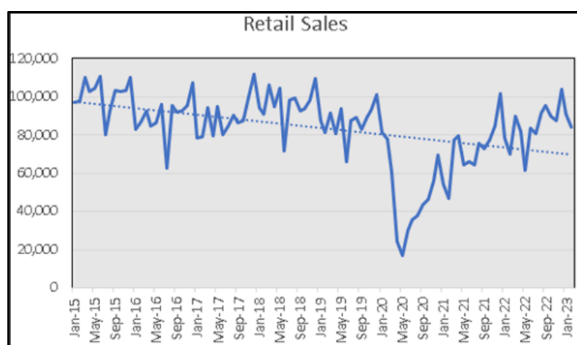
ABSTRAK

Persaingan di industri ini semakin meningkat seiring dengan masuknya peserta usaha baru. GAIKINDO sebagai organisasi otomotif di Indonesia, mengalami penurunan penjualan pada tahun 2020 yang disebabkan oleh pandemi *covid-19*. Oleh karena itu, untuk bersaing dengan pelaku industri otomotif lainnya, diperlukan optimasi dalam perencanaan yang dilakukan oleh GAIKINDO agar hasilnya menjadi lebih optimal. Penelitian ini akan menguji metode *Time Series* menggunakan algoritma SARIMA dengan melakukan pendekatan dengan metodologi CRISP-DM. Pengujian ini akan dilakukan menggunakan *Google Colaboratory* dan *dataset* yang digunakan adalah data *Retail Sales* dari Januari 2015 hingga Februari 2023. Hasil dari metode peramalan menunjukkan tingkat akurasi model yang terbaik, dengan akurasi parameter AR (p) = 1, *difference* (d) = 1, MA (q) = 1, AR musiman (P) = 1, *difference* musiman (D) = 0, MA musiman (Q) = 0, dan periode (s) = 12 atau SARIMA(1, 1, 1)(1, 0, 0)¹² menghasilkan nilai MAPE sebesar 0,407%. Dengan demikian, hasil prediksi *Retail Sales* pada GAIKINDO yang memungkinkan tidak akan jauh dari hasil *forecast* yang diperoleh nanti.

Kata kunci: *Time Series*, SARIMA, CRISP-DM.

1. PENDAHULUAN

Industri otomotif memiliki peran penting dalam perekonomian Indonesia, sehingga dengan semakin kompetitifnya industri otomotif di Indonesia membuat semua perusahaan berusaha bersaing untuk menjadi yang terbaik dalam pemasaran produk mereka. [1]. GAIKINDO (Gabungan Industri Kendaraan Bermotor Indonesia) adalah sebuah organisasi yang beranggotakan perusahaan-perusahaan otomotif di Indonesia. [2].



Gambar 1. Grafik *Retail Sales* Gaikindo Januari 2015 – Februari 2023

Pada Gambar 1. dari Januari 2015 hingga Februari 2023 memiliki data anomali di Januari 2020 hingga Mei 2020, hal tersebut bertepatan dengan adanya *covid-19* atau pandemi yang menyebabkan penjualan kendaraan menjadi menurun, sehingga membuat pola penjualan nya menjadi tidak stabil [3]. Namun, setelah bulan Mei 2020 sedikit demi sedikit data penjualan pada GAIKINDO sudah mulai kembali membaik seperti sebelum *covid-19* atau pandemi hingga Februari 2023.

Walaupun data penjualan di GAIKINDO sudah mulai meningkat kembali, tetapi dampak yang diberikan oleh kejadian tersebut membuat *trendline* pada data penjualan GAIKINDO menjadi sangat turun drastis. Oleh karena itu pada penelitian kali ini adalah untuk mencari tahu hasil peramalan atau *forecast* penjualan pada GAIKINDO selama beberapa tahun ke depan dengan cara melihat data historis penjualan pada GAIKINDO. Sehingga pihak GAIKINDO bisa memberikan informasi atau urgensi pada tiap-tiap perusahaan yang terlibat dalam organisasi tersebut. Hasil dari prediksi tersebut juga akan memberikan bentuk pola atau tren yang membantu dalam mengidentifikasi data tersebut.

Berdasarkan keterangan diatas, peneliti mengajukan metode *Time Series* dengan algoritma SARIMA (*Seasonal Autoregressive Integrated Moving Average*) untuk melakukan uji peramalan atau *forecast*. Dengan metode *Time Series* diharapkan dapat membantu mengidentifikasi anomali pada tren, musiman, dan pola lainnya dalam data penjualan selama beberapa tahun kedepan [4].

2. TINJAUAN PUSTAKA

2.1. Peramalan(Forecasting)

Forecasting secara umum sebagai proses memperkirakan nilai masa depan dari suatu variabel atau deret waktu berdasarkan data historis dan informasi yang tersedia. Peramalan adalah metode untuk memperkirakan suatu nilai di masa depan dengan menggunakan data masa lalu [5]. peramalan dapat diartikan sebagai sebuah prediksi atau estimasi mengenai peristiwa yang akan terjadi di masa depan, yang bertujuan untuk melakukan perencanaan dan pengendalian operasional yang lebih baik.

2.2. Deret Waktu (Time Series)

Deret waktu atau *time series* merupakan sekumpulan data atau informasi yang terkumpul secara berkala atau dalam interval waktu tertentu. Deret waktu adalah suatu rangkaian observasi yang tercatat pada interval waktu yang teratur, seringkali, dalam waktu singkat seperti jam, hari, atau minggu [6]. Adapun beberapa jenis pola pada data deret waktu, yaitu ada pola *Trend*, *Seasonality*, *Cyclical*, dan *Random Variations*. Pada pola *trend*, terdapat 3 jenis bentuk *trend* yaitu ada *uptrend*, *downtrend*, dan *sideway trend*.

2.3. CRISP-DM

CRISP-DM sebagai suatu proses standar dalam pengembangan proyek *data mining* yang terdiri dari enam tahapan yang terstruktur, dimulai dari pemahaman bisnis hingga implementasi hasil analisis. Dari uraian di atas, dapat disimpulkan bahwa ada enam tahapan yaitu: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* [7]. Dalam tahap *Business Understanding* memungkinkan untuk menentukan persoalan bisnis yang ingin diselesaikan, mengevaluasi kelayakan dan mengidentifikasi sumber daya yang dibutuhkan. *Data Understanding* proses data yang diperoleh dieksplorasi dan dianalisis untuk memahami karakteristik data, kesenjangan atau kelemahan data, dan mengevaluasi kualitas data. *Data Preparation*, tahap ini melibatkan pembersihan data, pemilihan atribut, dan penggabungan data yang berbeda. Tahap *Modeling*, tahap model statistik atau algoritma *machine learning* yang digunakan untuk membangun model prediksi. *Evaluation*, model yang dibangun dievaluasi dan dianalisis untuk memastikan bahwa model tersebut memenuhi tujuan proyek dan memberikan prediksi yang akurat. *Deployment*, proses ini memastikan bahwa performa model peramalan dapat optimal dan konsisten sehingga menghasilkan prediksi yang akurat.

2.4. Stationeritas

Stationeritas merujuk pada sifat atau karakteristik dari sebuah proses stokastik yang tidak berubah seiring waktu. Stationeritas sebagai suatu kondisi di mana statistik data seperti *mean*, *variance*, dan *covariance* tidak berubah secara signifikan dari waktu ke waktu [8]. Jika data yang dimiliki menandakan tidak stationer tetapi langsung masuk ke tahap *modeling*, maka hasil prediksi atau evaluasi nya menjadi kurang baik.

2.5. Uji Rolling Statistic

Rolling statistic merupakan sebuah metode analisis data yang umumnya digunakan dalam analisis *time series* untuk mengamati dan menggambarkan nilai statistik tertentu, misalnya rata-rata atau deviasi standar, pada suatu *dataset* dengan menghitung nilai tersebut dalam jendela waktu yang terus bergeser. Tujuan dari metode ini adalah untuk memberikan visualisasi yang memudahkan identifikasi

karakteristik kestasioneran data. Dengan menerapkan jendela waktu yang bergeser secara terus-menerus, dapat mengetahui perubahan nilai statistik dari waktu ke waktu, yang membantu dalam memahami tren dan pola dalam *dataset time series*. *Rolling statistic* sebagai suatu teknik analisis data yang berguna untuk mengidentifikasi pola musiman dan tren dalam data dengan menghitung nilai statistik tertentu (seperti rata-rata atau deviasi standar) dalam suatu data dengan mengambil nilai tersebut dalam jendela waktu yang bergeser secara terus-menerus [9].

2.6. Uji ADF

ADF (*Augmented Dickey Fuller*) merupakan sebuah metode uji akar unit (*unit root test*) yang digunakan untuk menentukan apakah suatu deret waktu stasioner atau tidak stasioner. Jadi dengan menggunakan ADF dapat membantu dalam menstationerkan data yang dimiliki dengan memperhatikan nilai dari *p-value* dan tingkat keyakinan 1%, 5%, dan 10%. ADF sebagai suatu metode yang memperluas uji *Dickey-Fuller* untuk mengatasi kasus data yang memiliki ketergantungan autokorelasi yang tinggi dan non-stasioneritas dengan mengasumsikan adanya tren deterministik pada data [10].

2.7. Fungsi ACF

ACF (*autocorrelation function*) adalah sebuah fungsi dalam analisis deret waktu yang digunakan untuk menunjukkan seberapa kuat ketergantungan antara nilai-nilai dalam deret waktu dengan nilai-nilai sebelumnya (autokorelasi). ACF sebagai suatu fungsi yang menunjukkan korelasi antara nilai-nilai data pada waktu yang sama dengan nilai-nilai data pada waktu sebelumnya, serta digunakan untuk mengidentifikasi struktur ketergantungan waktu dalam data [11]. Dalam ACF, korelasi dihitung dengan menggunakan koefisien korelasi *Pearson* antara deret waktu dengan dirinya sendiri yang telah diubah sedemikian rupa (misalnya dengan menghilangkan tren atau musiman). Koefisien korelasi memiliki rentang nilai antara -1 hingga 1. Jika nilainya adalah 1, maka menunjukkan adanya korelasi yang sempurna antara nilai-nilai dalam deret waktu dengan nilai-nilai sebelumnya. *Output* dari ACF ini yaitu nilai *autoregressive* (AR), *differencing*, dan *moving average* (MA).

2.8. Uji PACF

PACF (*Partial autocorrelation function*) adalah suatu fungsi matematis yang digunakan dalam analisis data *time series* untuk menunjukkan hubungan antara observasi dalam suatu data dan observasi sebelumnya dalam data tersebut, dengan menghilangkan pengaruh observasi pada interval waktu yang lebih jauh. PACF sebagai suatu fungsi yang menunjukkan korelasi antara nilai-nilai data pada waktu yang sama dengan nilai-nilai data pada waktu sebelumnya, setelah menghilangkan pengaruh observasi pada interval waktu yang lebih jauh, serta digunakan untuk

mengidentifikasi struktur ketergantungan waktu dalam data [11]. *Output* dari PACF ini yaitu nilai musiman (S), *seasonal autoregressive* (SAR), *seasonal differencing*, dan *seasonal moving average* (SMA)

2.9. SARIMA

SARIMA sebagai suatu model *time series* yang terdiri dari komponen musiman, *autoregressive* (AR), integrasi, dan *moving average* [12]. SARIMA (*Seasonal AutoRegressive Integrated Moving Average*) adalah salah satu metode yang digunakan dalam analisis data *time series* untuk memodelkan data dengan komponen musiman. Model SARIMA merupakan pengembangan dari model ARIMA yang digunakan untuk memodelkan data *time series* musiman. Model ini dinotasikan dalam bentuk $(p, d, q)(P, D, Q)^s$ di mana p, d , dan q adalah parameter untuk komponen *non-musiman*. P, D , dan Q adalah parameter untuk komponen musiman, dan s adalah panjang periode musiman dalam data. Model SARIMA digunakan ketika data *time series* menunjukkan pola musiman yang perlu diperhitungkan dan data tersebut juga memiliki komponen tren dan fluktuasi yang perlu diatasi melalui *differencing*.

2.10. Model Evaluasi

MAPE (*Mean Absolute Percentage Error*), Metrik evaluasi ini mengukur rata-rata persentase kesalahan antara nilai aktual dan nilai prediksi. Semakin kecil nilai MAPE, semakin akurat model peramalan rendah.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - \hat{x}_t}{x_t} \right| \times 100\%$$

Diketahui:

- n : Banyaknya data
- x_t : peramalan ke $-t$
- $\hat{x}_t$: data ke $-t$

3. METODE PENELITIAN

Pada penelitian ini, metodologi yang digunakan adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari enam tahap dalam pengolahan data, yaitu: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*.



Gambar 2. Flow CRISP-DM

(https://www.researchgate.net/figure/CRISP-DM-Model-Taylor-2017_fig1_326235288)

4. HASIL DAN PEMBAHASAN

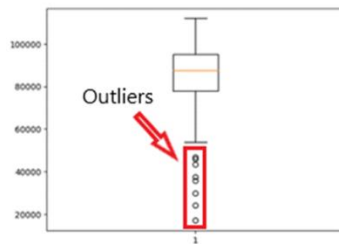
Pada tahap *business understanding*, hasil analisa ini menggunakan unsur 5W+1H dan menunjukkan bahwa terdapat anomali pada data penjualan di GAIKINDO yang menyebabkan penurunan penjualan dan produksi di sektor industri otomotif Indonesia. Anomali ini terlihat pada Gambar 1. yang menunjukkan *trendline* menurun. Peristiwa ini terjadi pada Januari 2020 hingga Mei 2020, di mana dampak pandemi *covid-19* mulai terasa di Indonesia. Pandemi ini telah mempengaruhi berbagai aspek kehidupan di Indonesia, termasuk kesehatan, ekonomi, sosial, psikologis, pendidikan, dan lain-lain. Berita tersebut dapat ditemukan di laman GAIKINDO (<https://www.gaikindo.or.id/wabah-resesi-ekonomi-dan-turunnya-penjualan-mobil-48-persen-pada-2020>). Untuk mengatasi masalah ini, solusi yang ditawarkan adalah menggunakan analisis deret waktu dengan algoritma SARIMA. Dengan menggunakan metode ini, diharapkan dapat mendeteksi perubahan penjualan di masa depan, baik itu tren atau pola penjualan. Hal ini akan membantu dalam pengambilan keputusan terkait stok atau produksi selama periode atau musim kedepannya.

Pada tahap *data understanding*, pertama melakukan proses pengumpulan data mentah atau *dataset* awal dengan menjadikan *Retail Sales* dan tanggal penjualan sebagai variabel. Lalu setelah itu pindah kan data mentah tersebut ke *Excel*. *Dataset* yang digunakan berasal dari data publik yang diambil dari *website* GAIKINDO (<https://files.gaikindo.or.id/>).

Tabel 1. Data Retail Sales Gaikindo Januari 2015 – Februari 2023

Bulan	Tahun								
	2015	2016	2017	2018	2019	2020	2021	2022	2023
Jan	97,067	82,676	78,297	94,120	87,555	81,059	53,997	78,568	90,872
Feb	97,426	87,057	79,127	90,852	81,293	77,865	46,941	69,989	83,973
Mar	110,118	92,510	94,435	106,050	91,115	60,440	77,514	89,947	-
Apr	102,920	84,524	79,621	94,727	80,622	24,275	79,502	81,616	-
May	104,103	86,339	94,856	104,153	93,881	17,083	64,175	61,558	-
Jun	110,335	95,965	80,143	71,778	65,750	29,858	65,715	83,579	-
Jul	79,909	62,427	84,735	98,182	87,611	35,799	64,033	80,520	-
Aug	92,684	95,227	90,343	99,310	89,258	37,654	75,820	91,163	-
Sep	103,426	91,815	86,563	92,355	82,853	43,357	72,637	95,426	-
Oct	102,386	92,646	87,623	93,632	88,914	46,131	77,002	89,651	-
Nov	103,401	95,252	99,690	98,002	93,318	56,102	84,544	87,674	-
Dec	110,191	107,109	111,963	109,480	100,847	69,139	101,468	103,891	-

Selanjutnya data tersebut dimasukkan ke dalam *Google Colaboratory* untuk di analisa, mulai dari mengecek jumlah data yang masuk, tipe data, *missing value*, dan *outliers*.

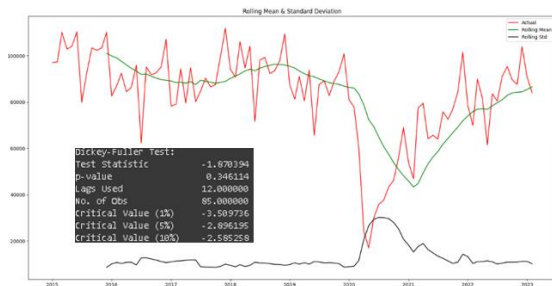


Gambar 3. Outliers

Ketika data yang masuk sudah sesuai, tipe data juga sudah sesuai dan tidak ada data yang kosong, maka selanjutnya mengecek *outliers*. Pada Gambar 3. Dijelaskan bahwa data yang dimiliki terdapat 8 data *outliers*, dan nilai tersebut ada: 17083, 24275, 29858, 35799, 37654, 43357, 46131, dan 46941. Setelah diteliti kembali data *outliers* tersebut merupakan data dari permasalahan kali ini yaitu disaat terjadinya pandemi *covid-19*, sehingga langkah yang diambil adalah dengan tidak menghapus data tersebut.

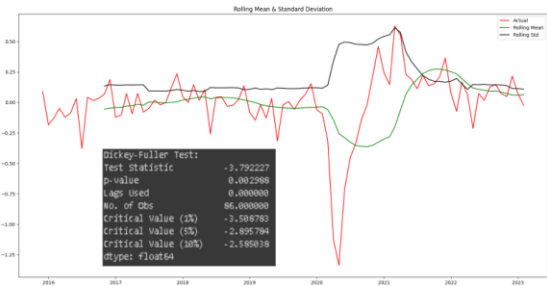
Pada tahap selanjutnya yaitu *data preparation*, langkah pertama adalah dengan melakukan penghapusan pada variabel yang tidak diperlukan. Setelah itu, melakukan *set index* pada variabel Tanggal.

Apabila data sudah dikatakan rapih dan sesuai, maka langkah selanjutnya adalah *modeling*. Tahap pertama yang perlu dilakukan adalah dengan mengecek ke stationeran data yang dimiliki, hal tersebut dapat dilakukan dengan menggunakan uji *Rolling Statistic*, hasil uji tersebut disampaikan pada Gambar 5.



Gambar 4. Uji Rolling Statistic

Pada Gambar 4. menjelaskan bahwasannya data *actual* dan *rolling (mean dan std)* memiliki perbedaan yang cukup signifikan, dalam hal ini menunjukkan bahwa seri data ini tidak bersifat stasioner. Oleh karena itu, dilanjutkan dengan melakukan uji ADF (*Augmented Dickey Fuller*) untuk menstationerkan data tersebut dengan melakukan *differencing* satu kali, hasil uji tersebut disampaikan pada Gambar 6.



Gambar 5. Uji ADF

Pada Gambar 5. dijelaskan bahwa hasil data *actual* dan *rolling (mean dan std)* memiliki perbedaan yang tidak cukup signifikan dikarenakan hasil *Dickey-Fuller test* ini untuk tingkat keyakinan 1%, 5%, dan 10% sudah menunjukkan *series* ini stasioner dimana $p\text{-value } 0.022235 < 0.05$. Dengan demikian, kita dapat menolak hipotesis nol (H_0) dan menerima hipotesis alternatif (H_1), yang menyatakan bahwa seri data ini bersifat stasioner. Setelah data di katakan stationer, maka selanjutnya membagi data *train test* nya menjadi 60:40, di mana 40% dari 98 akan menjadi 39.2, yang dibulatkan menjadi 39. Dengan demikian, nilai *train* akan menjadi $98 - 39 = 59$, sementara nilai *test* akan menjadi sisa dari nilai *train*, yaitu 41. Hasil tersebut disampaikan pada Gambar 6.

```
train = data[0:int(len(data)*0.59)]
test = data[int(len(data)*0.59):]
```

Gambar 6. Nilai train dan test

Setelah membagi data *train test* nya, maka selanjutnya mencari nilai *acf* dan *pacf* nya menggunakan konsep algoritma *tuning*, hasil tersebut disampaikan pada Gambar 7.

```
for param in pdq: #Non-Seasonal
    for param_seasonal in seasonal_pdq:
        try:
            temp_model = sm.tsa.statespace.SARIMAX(train_log,
                                                    order = param,
                                                    seasonal_order = param_seasonal)
            results = temp_model.fit()

            # print("SARIMAX[%d][%d] - AIC:{}".format(param, param_seasonal, results.aic))
            if results.aic < best_aic:
                best_aic = results.aic
                best_pdq = param
                best_seasonal_pdq = param_seasonal #store the best param values : AIC, P, D, Q (Non-S / Seas)
        except:
            #print("Unexpected error:", sys.exc_info()[0])
            continue
    print("Best SARIMAX[%d][%d] model - AIC:{}".format(best_pdq, best_seasonal_pdq, best_aic))

Best SARIMAX(1, 1, 1)x(1, 0, 0, 12)12 model - AIC:-187.9244881426892
```

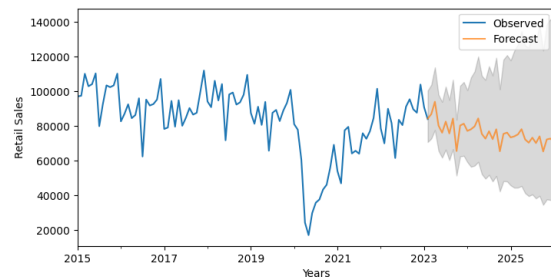
Gambar 7. Best SARIMA

Pada Gambar 7. terlihat adanya pola dengan menggunakan model ARIMA musiman orde (1, 1, 1)x(1, 0, 0, 12). Hal ini menunjukkan bahwa terdapat komponen *autoregressive* (AR) orde 1, komponen *moving average* (MA) orde 1, *differencing* orde 1, serta komponen *autoregressive musiman* (SAR) orde 1, komponen *moving average musiman* (SMA) orde 0, *differencing musiman* orde 0, periode musiman 12, dan nilai *Akaike Information Criterion* (AIC) adalah -187.93.



Gambar 8. Ekstrak nilai prediksi dan nilai kebenaran dari deret waktu

Setelahnya, terlihat pada Gambar 8. bahwa data pada uji *test* memiliki kemiripan dengan data *forecast*.



Gambar 9. Forecast 3 tahun ke depan

Pada Gambar 9. ditampilkan hasil *forecast* penjualan kendaraan (*Retail Sales*) untuk periode 3 tahun ke depan, beserta dengan nilai kepercayaan 95% dan 99%. Untuk *detail* nilai nya bisa dilihat pada Tabel 2.

Tabel 2. Data *forecast* Februari 2023 - Januari 2026

Tanggal	Forecast	lower_ci_95	upper_ci_95	lower_ci_99	upper_ci_99
Feb 2023	84285.02	70741.7	100421.17	70741.7	100421.2
Mar 2023	86854.37	72732.28	103718.46	72732.28	103718.5
Apr 2023	93981.58	77615.8	113798.17	77615.8	113798.2
May 2023	80244.32	65687.37	98027.234	65687.37	98027.23
Jun 2023	76151.25	61744.17	93920.003	61744.17	93920
Jul 2023	82538.42	66335.02	102699.76	66335.02	102699.8
Aug 2023	75705.9	60324.66	95008.955	60324.66	95008.95
Sep 2023	84300.58	66622.43	106669.59	66622.43	106669.6
Oct 2023	65551.65	51394.43	83608.649	51394.43	83608.65
Nov 2023	80284.02	62461.81	103191.44	62461.81	103191.4
Dec 2023	81347.05	62817.37	105342.56	62817.37	105342.6
Jan 2024	77179.48	59167.57	100674.61	59167.57	100674.6
Feb 2024	78119.24	56496.22	108018.12	56496.22	108018.1
Mar 2024	79793.71	57124.4	111459.11	57124.4	111459.1
Apr 2024	84364.38	59298.5	120025.78	59298.5	120025.8
May 2024	75455.17	52256.39	108952.85	52256.39	108952.9
Jun 2024	72716.06	49611.46	106580.73	49611.46	106580.7
Jul 2024	76972.4	51775.31	114431.96	51775.31	114432
Aug 2024	72415.45	48044.38	109149.04	48044.38	109149
Sep 2024	78129.42	51150.62	119337.87	51150.62	119337.9
Oct 2024	65411.92	42275.81	101209.62	42275.81	101209.6
Nov 2024	75481.53	48176.58	118262.06	48176.58	118262.1
Dec 2024	76186.04	48037.27	120829.35	48037.27	120829.4
Jan 2025	73408.14	45739.47	117814.1	45739.47	117814.1
Feb 2025	74038.31	44423.57	123395.57	44423.57	123395.6
Mar 2025	75155.66	44476.77	126996.03	44476.77	126996
Apr 2025	78171.18	45407.92	134574.16	45407.92	134574.2
May 2025	72245.98	41281.69	126435.75	41281.69	126435.8
Jun 2025	70383.68	39562.87	125214.96	39562.87	125215
Jul 2025	73268.98	40537.81	132428.06	40537.81	132428.1
Aug 2025	70178.06	38232.2	128817.07	38232.2	128817.1
Sep 2025	74045.12	39735.93	137977.91	39735.93	137977.9
Oct 2025	65313.41	34538.34	123510.31	34538.34	123510.3
Nov 2025	72263.81	37668.37	138632.46	37668.37	138632.5
Dec 2025	72739.52	37386.76	141521.7	37386.76	141521.7
Jan 2026	70856.14	35920.54	139769.4	35920.54	139769.4

Setelah sudah di *forecast*, maka perlu melakukan uji evaluasi model untuk mengetahui kualitas dari hasil *forecast* tersebut.

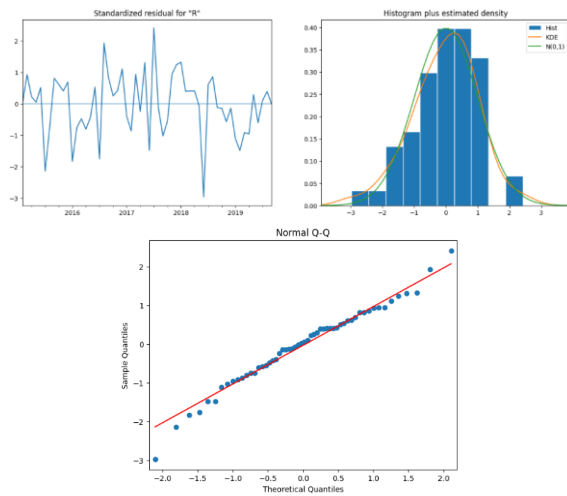
The Mean Absolute Percentage Error of our forecasts is 0.407 %

Gambar 10. Hasil MAPE

Pada Gambar 10. Di dapatkan MAPE sebesar 0.407%, hal tersebut merupakan hasil uji yang cukup baik karena masih mendekati angka nol.

Pada tahap *deployment* ini perlu dilakukan pengujian model, memastikan bahwa model

memberikan hasil yang akurat dan konsisten. Berikut disampaikan pada Gambar 11.



Gambar 11. Evaluation model

Pada Gambar 12. Dijelaskan bahwa grafik *standardized residuals* menunjukkan bahwa sebagian besar titik sudah berada pada $y=0$ yang berarti bahwa prediksi tidak menyimpang jauh dari data aktual, grafik histogram sudah menunjukkan kecenderungan normal (garis hijau adalah grafik garis dari ujung histogram) dan KDE (*Kernel Density Estimation*) menunjukkan grafik kecocokan prediksi, dan *Q-Q plot* adalah *probability plot*, yang mana digunakan untuk mengecek apakah data kita berada pada garis normal.

5. KESIMPULAN DAN SARAN

Berdasarkan semua tahapan penelitian yang telah dilakukan dalam menerapkan *data mining* untuk memprediksi penjualan masa depan GAIKINDO menggunakan algoritma SARIMA, dapat disimpulkan hasil *forecast* model SARIMA terbaik dengan parameter AR (p) = 1, *difference* (d) = 1, MA (q) = 1, AR musiman (P) = 1, *difference* musiman (D) = 0, MA musiman (Q) = 0, dan periode (s) = 12 dengan nilai AIC = -187.929. adapun saran untuk pengembangan penelitian yaitu diharapkan dapat menggunakan algoritma lain dan *tools* yang berbeda dengan variabel yang lebih banyak.

DAFTAR PUSTAKA

- [1] "Industri Otomotif Semakin Berdaya Saing," 25 Maret 2019. [Online]. Available: <https://www.indonesia.go.id/narasi/indonesia-dalam-angka/ekonomi/industri-otomotif-semakin-berdaya-saing>.
- [2] "Profile GAIKINDO," [Online]. Available: <https://www.gaikindo.or.id/profil/>.
- [3] "Tahun 2020: Wabah, Resesi Ekonomi, dan Turunnya Penjualan Mobil 48 Persen," 2021. [Online]. Available: <https://www.gaikindo.or.id/wabah-resesi-ekonomi-dan-turunnya-penjualan-mobil-48-persen-pada-2020/>.
- [4] M. A. Rohman, "Cara Memprediksi Penjualan Anda Menggunakan Ekonometrika Time Series," [Online]. Available: <http://sekolahstata.com/cara-memprediksi-penjualan-anda-menggunakan-ekonometrika-time-series/>.
- [5] K. M. Henry and E. Turban, Pengantar Teknologi Informasi, J. Haryanto and E. Ganis Sukoharsono, Eds., Andi, 2002.
- [6] C. Chatfield, The Analysis of Time Series: An Introduction, Sixth Edition (Chapman & Hall/CRC Texts in Statistical Science), 6th ed., Chapman and Hall/CRC, 2003.
- [7] R. Wirth and J. Hipp, CRISP-DM: towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, Springer-Verlag, 2000, pp. 29-39.
- [8] G. E. Box and G. M. Jenkins, Time Series Analysis: Forecasting and Control, 4th ed., Wiley, 2008.
- [9] E. W. Gilbert and M. A. Varughese, An Introduction to Time Series Analysis and Forecasting: With Applications of SAS® and SPSS®, John Wiley & Sons, 2020.
- [10] P. C. Phillips and P. Perron, Testing for a Unit Root in Time Series Regression, vol. 75, Biometrika, 1988, pp. 335-346.
- [11] M. J. Yar and C. Chatfield, The Analysis of Time Series: An Introduction, CRC Press, 2014.
- [12] G. Box and G. Jenkins, Time Series Analysis: Forecasting and Control, San Francisco: ACM Digital Library, 1976.