

PREDIKSI PENJUALAN BERBASIS TIME SERIES MENGGUNAKAN MODEL SARIMA PADA PT. AWI JAYA INDOTAMA

Wibisono

*Program Studi Teknik Informatika ,FTI, Institut Teknologi Budi Utomo Jakarta,
wibisono@itbu.ac.id*

Abstrak

Penelitian ini bertujuan untuk merancang sistem prediksi penjualan pipa HDPE yang lebih presisi guna menunjang pengambilan keputusan strategis di PT. Awi Jaya Indotama. Sebagai perusahaan yang bergerak di bidang distribusi pipa HDPE, PT. Awi Jaya Indotama menghadapi kendala dalam memperkirakan volume penjualan akibat dinamika pasar dan ketiadaan sistem prediksi berbasis data historis. Untuk menjawab permasalahan tersebut, digunakan pendekatan analisis deret waktu dengan model Seasonal Autoregressive Integrated Moving Average (SARIMA), serta kerangka kerja CRISP-DM sebagai metodologi analisis data. Data penjualan bulanan historis selama lebih dari satu dekade diolah melalui tahapan pra-pemrosesan, pemodelan, evaluasi, dan implementasi. Proses evaluasi dilakukan menggunakan indikator performa seperti MAE, MSE, RMSE, dan MAPE guna menilai tingkat akurasi hasil prediksi. Temuan penelitian menunjukkan bahwa model SARIMA mampu mengidentifikasi pola musiman dan tren historis secara signifikan, serta menghasilkan proyeksi yang relevan untuk mendukung perencanaan stok dan distribusi. Sebagai luaran, dikembangkan aplikasi prediksi berbasis Flask yang menyajikan hasil peramalan secara interaktif. Dengan demikian, penelitian ini memberikan kontribusi praktis dalam penerapan teknik data mining di sektor distribusi industri serta memperkaya kajian ilmiah mengenai penggunaan model SARIMA dalam prediksi penjualan.

Kata Kunci : Prediksi Penjualan, Time Series, SARIMA, CRISP-DM, Pipa HDPE, Flask.

1. PENDAHULUAN

Pipa HDPE dikenal luas karena keunggulannya dalam hal ketahanan terhadap korosi, fleksibilitas, kekuatan mekanik, serta daya tahan jangka panjang. Sebagai material yang banyak digunakan dalam proyek air bersih, jaringan gas, hingga saluran limbah. Permintaan terhadap pipa HDPE cenderung mengalami fluktuasi yang dipengaruhi oleh kondisi pasar, proyek pembangunan nasional, serta dinamika ekonomi secara umum. (R. A. Pramono, 2021)

PT. Awi Jaya Indotama adalah distributor di bidang perpipaan khususnya perpipaan High Density Polyethylene (HDPE) atau selang hitam beserta aksesorisnya. PT. Awi Jaya Indotama harus mampu mengantisipasi perubahan pasar tersebut secara cepat dan akurat. Dalam praktiknya, perusahaan menghadapi tantangan dalam perencanaan produksi dan distribusi akibat ketidakpastian volume penjualan bulanan. Kesalahan dalam memperkirakan jumlah kebutuhan pasar dapat mengakibatkan overstocking yang

membebani biaya gudang, atau sebaliknya. Terjadinya kekurangan stok yang berdampak pada kehilangan peluang penjualan dan menurunnya tingkat kepuasan pelanggan. Oleh karena itu, perusahaan membutuhkan suatu sistem prediksi yang andal untuk membantu memperkirakan penjualan dimasa mendatang secara lebih akurat.

Seasonal Autoregressive Integrated Moving Average (SARIMA) merupakan model statistik yang digunakan untuk prediksi pada data deret waktu (Time Series) yang mengandung unsur musiman. Model ini merupakan perluasan dari ARIMA dengan menambahkan komponen musiman guna menangkap pola berulang dalam periode tertentu, seperti bulanan atau tahunan (T. H. Noor. 2022).

2. METODOLOGI PENELITIAN

2.1. Jenis Penelitian

Jenis penelitian yang digunakan Penelitian ini merupakan penelitian kuantitatif dengan pendekatan deskriptif. Metode yang digunakan adalah analisis

deret waktu (time series forecasting) menggunakan model Seasonal Autoregressive Integrated Moving Average (SARIMA). Pendekatan kuantitatif dipilih karena penelitian ini berfokus pada data numerik historis untuk memodelkan dan meramalkan harga penjualan pipa HDPE. Pendekatan deskriptif digunakan untuk menggambarkan karakteristik data serta mengevaluasi kinerja model prediksi secara sistematis.

2.2 Metode Analisis Data

Penelitian ini menggambarkan alur pemikiran dalam melakukan analisis data hingga menghasilkan model prediksi yang akurat. Penelitian ini mengikuti pendekatan CRISP-DM (Cross-Industry Standard Process for Data Mining) yang terdiri atas enam tahapan utama seperti pada gambar yang ada dibawah ini :



Gambar 1 Fase CRISP-DM

Sumber : Penelitian Mandiri 2025

Model CRISP-DM berisikan daur hidup kegiatan data mining yang terdiri atas enam fase. Gambar 3.1 diatas menunjukkan keseluruhan tahapan pada model CRISP-DM. Salah satu keuntungan dari model ini adalah, tahapan fase dari model bukanlah sebuah tahapan yang kaku. Perpindahan maju dan mundur antara tiap fase yang berbeda bisa selalu dilakukan. Hal ini sesuai dengan sifat alami dari data mining itu sendiri, dimana proses data mining tidak selesai saat sebuah hasil di temukan, sebab proses data mining merupakan sebuah proses pembelajaran terus menerus. Berikut ini akan dijelaskan secara singkat apa saja yang dilakukan pada tiap fase dari model ini.

1) Fase *Business Understanding*

2) Fase *Data Understanding*

3) Fase *Data Preparation*

4) Fase *Modelling*

5) Fase Evaluasi

6) Fase *Implementation*

2.2 Skema Pelaksanaan Penelitian

Untuk mempermudah pemahaman mengenai tahapan-tahapan yang dilakukan dalam penelitian ini, maka disusun sebuah skema pelaksanaan penelitian berdasarkan pendekatan Cross Industry Standard Process for Data Mining (CRISP-DM). Skema ini menggambarkan alur proses yang dimulai dari pemahaman bisnis hingga evaluasi dan hasil akhir penelitian. Setiap tahapan dalam skema ini merepresentasikan proses yang saling terkait dan dilakukan secara sistematis guna mencapai tujuan dari penelitian. Skema pelaksanaan penelitian dapat dilihat pada Gambar berikut :



Gambar 2 Skema Proses CRISP-DM

Sumber : Penelitian Mandiri 2025

3. HASIL DAN PEMBAHASAN

3.1. *Business Understanding*

1. Tujuan Bisnis

Penelitian ini bertujuan untuk memprediksi penjualan produk pipa HDPE pada PT. Awi Jaya Indotama menggunakan pendekatan time series dengan model SARIMA. Dengan menerapkan metode CRISP-DM sebagai kerangka kerja analisis data, penelitian ini bertujuan untuk membantu perusahaan dalam memahami pola historis penjualan, mengidentifikasi tren, dan menghasilkan prediksi yang dapat digunakan sebagai dasar dalam pengambilan keputusan jstrategis, seperti perencanaan stok dan distribusi.

2. Permasalahan Bisnis

Fokus utama dari penelitian ini adalah bagaimana membangun model prediksi penjualan yang akurat berdasarkan data historis harian yang tersedia, dengan memanfaatkan model SARIMA sebagai metode time series. Permasalahan yang diangkat mencakup ketidakpastian permintaan yang mempengaruhi efisiensi rantai pasok, serta kurangnya sistem prediktif yang dapat memberikan wawasan kuantitatif kepada manajemen dalam merencanakan kebutuhan operasional secara proaktif.

3. Solusi yang Ditawarkan

Dalam menjawab permasalahan tersebut, peneliti membangun sebuah sistem prediksi penjualan berbasis model time series SARIMA, yang dapat menganalisis pola penjualan dari data historis dan memprediksi jumlah penjualan pada periode mendatang. Kemudian dibuat dalam bentuk aplikasi sederhana berbasis web agar mudah digunakan.

3.2. Data Understanding

Pada tahap data understanding bertujuan untuk memperoleh pemahaman mendalam terhadap data yang tersedia.

1. Load Data



Month	Sales_qty
2015-01-01	112
2015-02-01	118
2015-03-01	122
2015-04-01	129
2015-05-01	121

Gambar 3 Proses Load Data

Sumber : Penelitian Mandiri 2025

Pada gambar 4 menunjukkan proses Proses ini bertujuan untuk mengimpor dataset penjualan pipa HDPE dan mengonversi kolom waktu menjadi format datetime, sehingga siap digunakan dalam analisis deret waktu. Kolom "Month" dijadikan indeks karena metode prediksi SARIMA mensyaratkan data harus dalam

bentuk time series yang terindeks waktu secara historis.

Berdasarkan data awal, terlihat adanya variasi jumlah penjualan dari bulan ke bulan. Sebagai contoh, terdapat peningkatan yang cukup signifikan dari bulan Januari hingga Maret (112 menjadi 132), diikuti oleh penurunan ringan pada bulan April dan Mei. Pola fluktuasi ini dapat memberikan indikasi awal mengenai adanya kecenderungan musiman atau tren, yang selanjutnya akan dianalisis lebih mendalam menggunakan metode peramalan deret waktu seperti SARIMA.

2. Melihat Info df



```
# Melihat Info df
df.info()
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 124 entries, 2015-01-01 to 2025-04-01
Data columns (total 1 columns):
 #   Column      Non-Null Count  Dtype
---  ---
 0   Sales_qty   124 non-null     int64
dtypes: int64(1)
memory usage: 1.9 KB
```

Gambar 4 Cek Info df

Sumber : Penelitian Mandiri 2025

Pada gambar 5 memperlihatkan bahwa dataset terdiri atas 124 data penjualan bulanan dengan indeks waktu yang dimulai dari Januari 2015 hingga April 2025. Kolom utama yang digunakan, yaitu Sales_qty, bertipe bilangan bulat (int64) dan tidak mengandung nilai kosong, sehingga data tersebut memenuhi syarat untuk dianalisis sebagai deret waktu. Adanya indeks waktu bertipe DatetimeIndex menunjukkan bahwa struktur dataset telah sesuai dengan kebutuhan metode peramalan seperti SARIMA, yang memerlukan urutan waktu yang konsisten dalam analisisnya.

3. Menghitung Describe



```
# Menghitung Info df
df.describe()
```

	Sales_qty
count	124.000000
mean	120.500000
std	96.300000
min	104.000000
25%	117.500000
50%	120.000000
75%	121.000000
max	132.000000

Gambar 5 Menghitung Describe

Sumber : Penelitian Mandiri 2025

Dari gambar 5, nilai rata-rata penjualan per bulan adalah sekitar 250 unit, dengan simpangan baku sebesar 96.31, yang menunjukkan tingkat variasi atau penyebaran data penjualan cukup besar dari nilai rata-rata. Nilai minimum dan maksimum berturut-turut adalah 104 dan 505, menunjukkan adanya fluktuasi yang signifikan antar periode. Distribusi data juga menunjukkan bahwa 50% data berada di bawah nilai 235 (median), dan 75% data berada di bawah nilai 317.25.

4. Mengecek Indeks df

```
# Cek Indeks df
df.index

DatetimeIndex(['2015-01-01', '2015-02-01', '2015-03-01', '2015-04-01',
                '2015-05-01', '2015-06-01', '2015-07-01', '2015-08-01',
                '2015-09-01', '2015-10-01',
                ...,
                '2024-07-01', '2024-08-01', '2024-09-01', '2024-10-01',
                '2024-11-01', '2024-12-01', '2025-01-01', '2025-02-01',
                '2025-03-01', '2025-04-01'],
              dtype='datetime64[ns]', name='Month', length=124, freq='MS')
```

Gambar 6Cek Indeks df

Sumber : Penelitian Mandiri 2025

Pada Gambar 7 diperlihatkan bahwa struktur data dalam DataFrame telah menggunakan indeks bertipe waktu (DatetimeIndex) yang disusun secara bulanan, dimulai dari Januari 2015 hingga April 2025. Indeks tersebut terdiri atas 124 titik waktu yang merepresentasikan 124 bulan data historis. Setiap elemen indeks menunjukkan awal bulan (seperti: 2015-01-01, 2015-02-01, dan seterusnya), dengan label indeks diberi nama "Month", yang menegaskan bahwa interval waktu utama yang digunakan dalam analisis adalah per bulan.

5. Merubah Indeks df1 menjadi 'MS'

```
DatetimeIndex(['2015-01-01', '2015-02-01', '2015-03-01', '2015-04-01',
                '2015-05-01', '2015-06-01', '2015-07-01', '2015-08-01',
                '2015-09-01', '2015-10-01',
                ...,
                '2024-07-01', '2024-08-01', '2024-09-01', '2024-10-01',
                '2024-11-01', '2024-12-01', '2025-01-01', '2025-02-01',
                '2025-03-01', '2025-04-01'],
              dtype='datetime64[ns]', name='Month', length=124, freq='MS')
```

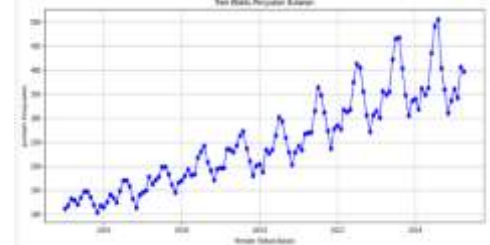
Gambar 7 Merubah Indeks df1 menjadi MS

Sumber : Penelitian Mandiri 2025

Penetapan `df.index.freq = 'MS'` menjadikan DataFrame lebih terstruktur dan sesuai untuk dianalisis sebagai data deret waktu

dengan interval bulanan. Langkah ini merupakan praktik yang disarankan sebelum dilakukan transformasi data berbasis waktu, pemodelan prediktif, maupun visualisasi yang mengandalkan dimensi temporal.

6. Menyvisualisasikan Tren Waktu

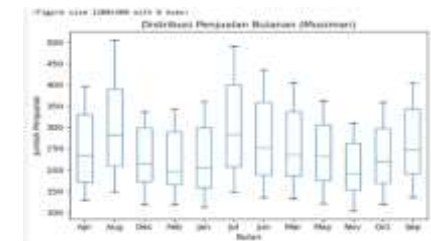


Gambar 8 Grafik Visualisasi Tren Waktu

Sumber : Penelitian Mandiri 2025

Dari visualisasi tersebut, terlihat Pola naik-turun yang tajam pada beberapa titik juga mengindikasikan adanya volatilitas atau faktor eksternal yang mempengaruhi volume penjualan dalam jangka pendek. Visualisasi ini sangat penting sebagai langkah awal dalam analisis deret waktu, khususnya dalam mengidentifikasi komponen tren, musiman, dan residu yang dapat dimodelkan lebih lanjut dengan pendekatan ARIMA atau SARIMA

7. Visualisasi Boxplot Musiman per Bulan



Gambar 10 Hasil Visualisasi Boxplot Boxplot Sumber : Penelitian Mandiri 2025

Beberapa temuan penting dari visualisasi ini:

- 1) Bulan Juli dan Agustus menunjukkan distribusi penjualan tertinggi dengan median yang tinggi serta rentang interkuartil yang lebar, menandakan adanya lonjakan

penjualan musiman pada pertengahan tahun.

- 2) Bulan November dan Februari menunjukkan median yang lebih rendah dibandingkan bulan lainnya, yang menunjukkan bahwa penjualan cenderung lebih rendah di akhir dan awal tahun.
- 3) Bulan Maret hingga Juni memiliki distribusi yang cukup stabil, tanpa fluktuasi ekstrem.
- 4) Hampir setiap bulan menunjukkan adanya outlier (nilai ekstrim), mengindikasikan beberapa periode dengan penjualan yang tidak biasa (sangat tinggi atau rendah).

Visualisasi ini mendukung adanya pola musiman (seasonality) yang kuat, yang menjadi pertimbangan penting dalam pemilihan model deret waktu, misalnya menggunakan SARIMA untuk memodelkan faktor musiman secara eksplisit.

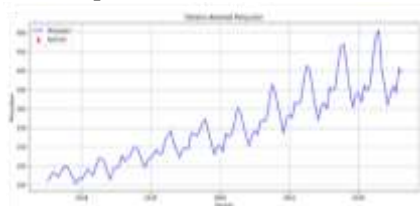
8. Mendeteksi Anomali

```
from sklearn.ensemble import IsolationForest
# Hitung z-score untuk setiap penjualan
df['z_score'] = zscore(df['Sales'])
# Tentukan ambang anomali (nilai z-score lebih dari 3 atau kurang dari -3)
df['anomaly'] = df['z_score'].apply(lambda x: abs(x) > 3)
# Tampilkan anomali
anomaly_idx = df[df['anomaly']]
print(anomaly_idx[['Sales', 'Date']])
print(anomaly_idx[['Sales', 'Date', 'z_score']])
```

Gambar 11 Mendeteksi Anomali
Sumber : Penelitian Mandiri 2025

Pada gambar 13 akan dilakukan mendeteksi anomali dengan menggunakan fungsi zscore() untuk menghitung skor Z dari data numerik, yaitu ukuran seberapa jauh suatu nilai rata-rata.

9. Menampilkan Anomali



Gambar 12 Menampilkan Anomali
Sumber : Penelitian Mandiri 2025

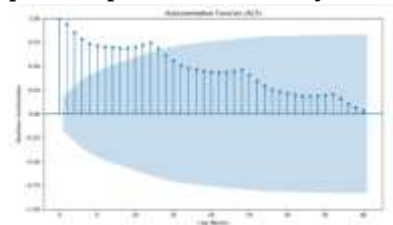
Beberapa hal yang dapat disimpulkan dari grafik ini:

- Secara umum, tren penjualan tetap menunjukkan kenaikan bertahap seiring waktu.
 - Dari hasil grafik tersebut tidak ditemukan adanya anomali.
- Dengan adanya deteksi anomali, perusahaan dapat mengevaluasi kembali strategi operasional atau pemasaran yang telah dilakukan pada periode tersebut.

10. ACF dan PCF

1) ACF

ACF (Auto Correlation Function) atau correlogram merupakan plot yang memberikan gambaran hubungan korelasi antara data time series dengan data itu sendiri di periode-periode sebelumnya



Gambar 13 Menampilkan Plot ACF
Sumber : Penelitian Mandiri 2025

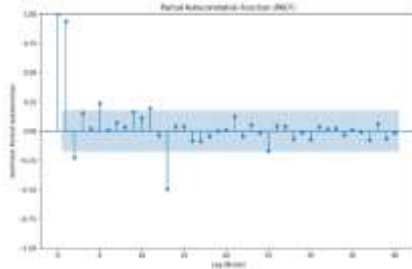
Grafik di atas menampilkan hasil analisis Autocorrelation Function (ACF) terhadap data penjualan bulanan. ACF digunakan untuk mengukur sejauh mana nilai penjualan pada bulan tertentu berkorelasi dengan nilai pada bulan-bulan sebelumnya. Sumbu X – Lag (Bulan) Sumbu horizontal menunjukkan nilai lag, yaitu jeda waktu dalam satuan bulan.

- Lag 1 mengacu pada korelasi antara bulan ini dan bulan sebelumnya.
 - Lag 12 mengacu pada korelasi antara bulan ini dan bulan yang sama di tahun sebelumnya.
- Sumbu Y – Koefisien Autokorelasi:

Sumbu vertikal menunjukkan nilai koefisien autokorelasi, yang berkisar dari -1 hingga 1:

- Nilai mendekati +1 menunjukkan korelasi positif yang kuat.
- Nilai 0 berarti tidak ada korelasi.
- Nilai mendekati -1 menunjukkan korelasi negatif yang kuat.

2) PACF



Gambar 14 Menampilkan Plot PACF

Sumber : Penelitian Mandiri 2025
Grafik di atas menampilkan hasil analisis Partial Autocorrelation Function (PACF) terhadap data penjualan bulanan hingga lag ke-40. SARIMA

Model SARIMA (Seasonal AutoRegressive Integrated Moving Average) merupakan pengembangan dari model ARIMA yang mampu menangani data deret waktu dengan pola musiman (seasonality). SARIMA terdiri dari dua bagian utama: komponen non-musiman (ARIMA) dan komponen musiman (seasonal).

3.3. Data Pre-Paration

1. Time Series Preprocessing

```
import pandas as pd

# Load data (file dummy)
df = pd.read_csv("data/ds_penjualan.csv", sep=";")

df['Month'] = pd.to_datetime(df['Month'], format='%m-%Y')
df.set_index('Month', inplace=True)
df.sort_index(inplace=True)

# Cek hasil
print(df.head())
```

Month	Sales_qty
2015-01-01	112
2015-02-01	118
2015-03-01	132
2015-04-01	129
2015-05-01	123

Gambar 15 Proses Awal Time Series

Sumber : Penelitian Mandiri 2025

Berdasarkan gambar yang ditampilkan, kita dapat melihat hasil dari proses awal data preparation yang ditujukan untuk analisis deret waktu (time series). Data yang

digunakan merupakan data penjualan bulanan yang disimpan dalam file CSV bernama ds_penjualan.csv. Langkah pertama yang dilakukan adalah membaca file tersebut ke dalam sebuah DataFrame menggunakan library pandas, dengan pemisah data berupa tanda titik koma (;), yang umum digunakan dalam format file regional Eropa atau Indonesia. Setelah data berhasil dimuat, kolom yang berisi informasi waktu, yaitu 'Month', dikonversi menjadi format datetime. Proses ini penting karena data deret waktu memerlukan pemahaman struktur tanggal agar bisa dianalisis secara kronologis. Format %m-%Y digunakan untuk membaca data dalam bentuk bulan dan tahun, misalnya "01-2015" menjadi 1 Januari 2015.

Selanjutnya, kolom 'Month' tersebut dijadikan sebagai index dari DataFrame. Hal ini dilakukan karena dalam analisis time series, urutan waktu berfungsi sebagai sumbu utama dalam pemodelan. Setelah itu, data diurutkan berdasarkan waktu agar model statistik seperti ARIMA atau SARIMA dapat membaca urutan data secara berurutan tanpa gangguan. Hasil akhirnya ditampilkan melalui fungsi print(df.head()), yang menunjukkan lima baris pertama dari data. Kita bisa melihat bahwa data telah tersusun rapi dengan indeks berupa tanggal di awal setiap bulan (misalnya 2015-01-01, 2015-02-01, dan seterusnya), serta kolom Sales_qty yang berisi jumlah penjualan per bulan. Contohnya, pada bulan Januari 2015 tercatat penjualan sebanyak 112 unit, dan pada bulan Maret 2015 meningkat menjadi 132 unit.

2. Proses Split Data menjadi Train dan Test Set

```

# Split data into train and test sets
train_data = df[:int(df.shape[0]*0.8)]
test_data = df[int(df.shape[0]*0.8):]

# Display the first few rows of the training data
print("Training Data (first 5 rows):")
print(train_data.head())

# Display the first few rows of the test data
print("Test Data (first 5 rows):")
print(test_data.head())

# Convert the date column to datetime format
train_data['date'] = pd.to_datetime(train_data['date'])
test_data['date'] = pd.to_datetime(test_data['date'])

# Sort the data by date
train_data = train_data.sort_values('date')
test_data = test_data.sort_values('date')

# Display the date range of the training and test data
print("Training Data Date Range: %s to %s" % (train_data['date'].min(), train_data['date'].max()))
print("Test Data Date Range: %s to %s" % (test_data['date'].min(), test_data['date'].max()))

```

Gambar 16 Split Data untuk Train Set dan Test Set

Sumber : Penelitian Mandiri 2025

Disini program menghitung berapa banyak data (baris) yang tersedia dalam DataFrame(df).

Sebanyak 80% dari data digunakan untuk pelatihan model. Perhitungan ini dibulatkan ke bawah agar hasilnya berupa angka bulat yang valid sebagai indeks

df_train berisi data dari awal hingga titik ke-80%.

df_test berisi sisa 20% data dari titik tersebut sampai akhir.

Karena ini data time series, maka urutannya tetap dijaga atau dengan kata lain tidak diacak.

Setelah data berhasil dibagi menjadi dua bagian berdasarkan indeks waktu, langkah selanjutnya adalah mencetak atau menampilkan rentang waktu dari masing-masing subset data, yaitu data latih (train set) dan data uji (test set). Proses ini dilakukan untuk memverifikasi bahwa pembagian data telah dilakukan dengan benar, serta memastikan bahwa data latih mencakup periode yang lebih lama (awal waktu) dan data uji mencakup periode yang lebih baru (masa depan). Hasil seperti gambar dibawah ini:

```

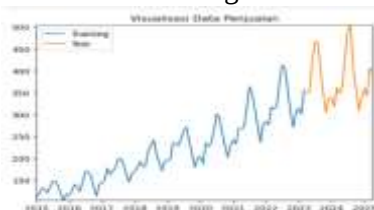
Train set: 2015-01-01 sampai 2023-03-01
Test set : 2023-04-01 sampai 2025-04-01

```

Gambar 17 Hasil Dari Proses Split Data

Sumber : Penelitian Mandiri 2025

3. Visualisasi Training dan Test Set



Gambar 18 Grafik Visualisasi Data Panjualan

Sumber : Penelitian Mandiri 2025

Grafik di atas memperlihatkan pembagian dataset penjualan bulanan menjadi dua bagian, yaitu:

- Data Training (berwarna biru): mencakup sekitar 80% awal dari total data, digunakan untuk melatih model prediksi.
- Data Testing (berwarna oranye): mencakup 20% akhir dari data, digunakan untuk mengukur performa model prediksi terhadap data yang belum pernah dilihat sebelumnya.

Rentang Waktu:

- Data Training: dari Januari 2015 hingga sekitar pertengahan 2023.
- Data Testing: dari pertengahan 2023 hingga April 2025.

Pola yang Terlihat:

- Terdapat tren naik yang konsisten dari tahun ke tahun, menunjukkan pertumbuhan jumlah penjualan.
- Terdapat juga fluktuasi musiman, yaitu kenaikan dan penurunan penjualan yang berulang secara periodik setiap tahun.
- Data testing mempertahankan karakteristik yang sama, yaitu tren meningkat dan pola musiman, sehingga cocok untuk dilakukan peramalan berbasis model SARIMA.

Tujuan Pembagian Data:

- Melatih model dengan data historis.
- Menguji seberapa baik model dapat memprediksi data masa depan.
- Menghindari overfitting, yaitu model terlalu cocok dengan data latih tetapi buruk dalam prediksi data baru.

Dengan pendekatan ini, performa model dapat diukur secara objektif menggunakan metrik evaluasi seperti MAE, RMSE, atau MAPE terhadap data testing.

3.4. Modeling

- Proses Ringkasan Model

SARIMAX Results			
Dep. Variable:		#	No. Observations:
Model:	SARIMAX(1, 0, 0)(0, 1, 0, 12)		Log Likelihood
Date:	Tu, 20 Jun 2023		AIC
Time:	12:12:46		BIC
Sample:	01-01-2015		HQIC
	- 04-01-2025		
Covariance Type:	opg		

Gambar19 Ringkasan Model SARIMA
Sumber : Penelitian Mandiri 2025

Tabel di atas menunjukkan ringkasan hasil pelatihan model SARIMA yang secara otomatis ditentukan menggunakan pendekatan Auto ARIMA, dengan konfigurasi sebagai berikut:

Kesimpulan:

Model SARIMA(1,0,0)(0,1,0,12)
yang dipilih oleh Auto ARIMA:

- Menangkap pola musiman dengan seasonal differencing 1 kali per tahun.
- Memiliki performa residual yang baik (tidak ada autokorelasi dan residual mendekati normal).
- Layak digunakan untuk melakukan peramalan penjualan pada data yang memiliki pola musiman dan tren jangka panjang.
- Proses Membuat Model Sarima

- Proses Membuat Model Sarima

```

# fit your model
print(model_FIT.summary())


```

(Unadjusted Results)

Dep. Variable:	Intercept	Age	Observations
model:	INTERCEPT, 1	AGE, 1	125
dvars:		AGE, 28	125
Train:		AGE, 28	125
Test:		AGE, 28	125
Sample:		AGE, 28	125
Estimation Type:	MLE		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	1.000	0.000	0.000	1.000	1.000	1.000
AGE	0.000	0.000	0.000	1.000	0.000	0.000

	variance	sd	skewness	kurtosis
Intercept	0.000	0.000	0.000	0.000
AGE	0.000	0.000	0.000	0.000

Gambar 20 Membuat Model Sarima
Sumber : Penelitian Mandiri 2025

Tabel di atas menampilkan ringkasan hasil pelatihan model SARIMA dengan konfigurasi:

- Model: SARIMA(0, 1, 0)(1, 0, 0, 12)
- 1. order = (0, 1, 0) → Tidak ada komponen AR maupun MA, namun dilakukan differencing satu kali (d=1) untuk menjadikan data stasioner.

2. seasonal_order = (1, 0, 0, 12)
→ Menambahkan komponen AR musiman (lag 12) dengan periode musiman tahunan (s=12), tanpa seasonal differencing atau MA musiman.

- Jumlah Observasi: 124 (dari Januari 2015 sampai April 2025)
- Log Likelihood: -415.934
- AIC: 835.867
- BIC: 841.286
- HQIC: 838.066

Nilai AIC dan BIC ini digunakan untuk membandingkan model — semakin rendah nilainya, semakin baik performa model dalam menyeimbangkan kompleksitas dan ketepatan

Parameter	Koefisien	Std. Err.	P-value	Interpretasi
α_0, α_1	1.1023	.0011	<.0001	Sangat signifikan, menunjukkan adanya pengaruh signifikan terhadap variabel yang diukur pada lag 12
σ^2	106.2388	13.164	<.0001	Varian residual model

Gambar 24 Koefisien Model
Sumber : Penelitian Mandiri 2025

- Nilai $P < 0.05$ menandakan parameter tersebut signifikan secara statistik.
- Koefisien AR musiman (lag 12) sebesar 1.10 mengindikasikan bahwa penjualan pada bulan tertentu sangat dipengaruhi oleh nilai penjualan setahun sebelumnya.

Uji Statistik	Nilai	P-value	Kesimpulan
Ljung-Box Q	10.43	0.30	Tidak ada autokorelasi yang signifikan pada residual
Jarque-Bera	0.82	0.66	Residual berdistribusi normal
Skew	-0.11		Residual seakut simetris
Kurtosis	3.36		Dekat dengan distribusi normal ($\text{ideal} = 3$)

Gambar 25 Diagnostik Resudial
Sumber : Penelitian Mandiri 2025

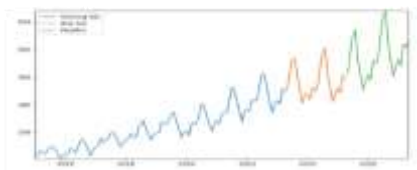
Diagnostik ini menunjukkan bahwa:

- Residual model tidak memiliki pola tertentu, mendekati white noise.
- Model sudah cukup baik, karena residual tidak autokorelasi dan terdistribusi mendekati normal.

Model SARIMA(0,1,0)(1,0,0,12) mampu:

- Menangkap pola musiman tahunan dengan sangat baik.
- Menyederhanakan model (tanpa komponen AR/MA biasa) tetapi tetap menghasilkan residual yang baik.
- Cocok untuk peramalan jangka pendek dan menengah, karena residual tidak menunjukkan masalah struktural.

4. Proses Visualisasi Sarima



Gambar 25 Script visualisasi sarima

Sumber : Penelitian Mandiri 2025

Gambar diatas merupakan script untuk membuat sebuah visualisasi sarima yang dimana akan menguji data frame training set, test set untuk menghasilkan prediksi. Grafik di atas menunjukkan hasil visualisasi prediksi model SARIMA terhadap data penjualan bulanan. Data dibagi menjadi tiga bagian yang direpresentasikan dengan warna berbeda:

- Training Set (biru): Data historis yang digunakan untuk melatih model, mencakup periode awal hingga sekitar pertengahan tahun 2023.
- Test Set (oranye): Data aktual dari sekitar pertengahan 2023 hingga awal 2025, digunakan untuk menguji akurasi model pada data yang tidak dilibatkan dalam pelatihan.
- Prediksi (hijau): Hasil prediksi penjualan untuk periode setelah April 2025, dihasilkan oleh model SARIMA berdasarkan pola musiman dan tren sebelumnya.

Interpretasi:

- Model SARIMA berhasil menangkap pola tren dan musiman dari data historis.
- Prediksi mengikuti pola naik-turun tahunan yang serupa dengan data sebelumnya, mencerminkan kemampuan model dalam mengenali komponen musiman (seasonality).
- Fluktuasi musiman yang signifikan terlihat jelas pada bagian prediksi, menunjukkan bahwa model memprediksi adanya siklus permintaan tahunan yang terus berulang.
- Perbedaan antara test set dan prediksi menunjukkan bahwa model memproyeksikan pertumbuhan penjualan yang masih cukup tinggi setelah 2025.

Kesimpulan:

- Visualisasi ini menunjukkan bahwa model SARIMA telah berhasil mengidentifikasi pola historis secara efektif dan mampu digunakan untuk prediksi jangka menengah.
- Performa model dapat dievaluasi lebih lanjut dengan membandingkan nilai prediksi pada test set dengan data aktual menggunakan metrik seperti MAE, RMSE, atau MAPE.

3.5. Evaluasi

1. Import Library evaluasi dan Hitung matrik evaluasi

Berdasarkan hasil evaluasi terhadap model prediksi penjualan menggunakan metrik statistik, diperoleh hasil sebagai berikut:

1. Mean Absolute Error (MAE) sebesar 95,71 menunjukkan bahwa rata-rata selisih absolut antara nilai aktual dan nilai prediksi adalah sekitar 96 unit penjualan. Ini berarti bahwa prediksi model memiliki deviasi rata-rata sebesar 96 unit dari

kenyataan, yang tergolong sedang tergantung pada skala penjualan harian atau bulanan.

2. Mean Squared Error (MSE) sebesar 12160,67 mengukur rata-rata dari kuadrat error, yang memberikan penalti lebih besar terhadap kesalahan besar. Nilai ini cukup tinggi, menandakan bahwa masih terdapat beberapa prediksi yang meleset jauh dari nilai aktualnya.
 3. Root Mean Squared Error (RMSE) sebesar 110,28 memberikan ukuran kesalahan yang sebanding satuan dengan data aslinya. RMSE yang lebih besar dari MAE menunjukkan bahwa terdapat sejumlah error besar (outlier) dalam prediksi yang perlu diperhatikan.
 4. Mean Absolute Percentage Error (MAPE) sebesar 0.259 atau 25.9%, menunjukkan bahwa rata-rata persentase kesalahan prediksi terhadap nilai aktual adalah sekitar 25.9%. Dalam konteks peramalan time series, nilai MAPE antara 20–30% tergolong akurasi sedang (reasonable forecasting accuracy), tetapi masih bisa ditingkatkan.
 5. Akurasi prediksi sebesar 74.13%, yang merupakan hasil dari $100 - \text{MAPE} (\%)$, menandakan bahwa secara umum model berhasil memprediksi tren penjualan dengan tingkat ketepatan yang cukup baik, meskipun masih menyisakan sekitar 25.9% tingkat kesalahan.
2. Perbandingan Data Aktual dan Prediksi



Gambar 26 Plot Perbandingan Data Aktual & Prediksi

Sumber : Penelitian Mandiri 2025

Grafik di atas menyajikan perbandingan antara data penjualan aktual dengan hasil prediksi model SARIMA untuk periode April 2023 hingga April 2025.

Dua garis dalam grafik merepresentasikan:

- Garis biru: Data penjualan aktual dari hasil observasi pada periode tersebut.
- Garis oranye putus-putus: Hasil prediksi dari model SARIMA berdasarkan pola historis data training.

Interpretasi Grafik:

- Secara umum, model SARIMA berhasil mengikuti tren musiman tahunan yang terlihat dalam data aktual.
- Pada periode tertentu, seperti pertengahan 2023 dan pertengahan 2024, prediksi model menunjukkan puncak penjualan yang cukup selaras dengan data aktual.
- Namun, terlihat adanya overestimasi pada beberapa titik, terutama saat puncak musiman di bulan-bulan tengah tahun, di mana prediksi melebihi angka aktual.
- Model juga sedikit menggeser waktu puncak dan lembah musiman, menandakan bahwa meskipun pola musiman dikenali, penyesuaian waktu (timing) tidak selalu tepat.

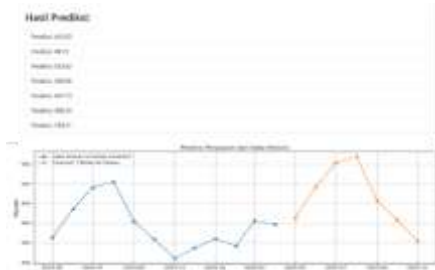
Evaluasi Model:

- Meskipun terdapat deviasi, model SARIMA menangkap arah pergerakan (naik/turun) dengan cukup baik.
- Kinerja model dapat ditingkatkan lebih lanjut melalui tuning parameter atau penambahan variabel eksternal (exogenous) jika tersedia.

- Performa model dapat dinilai secara kuantitatif menggunakan metrik seperti MAE, RMSE, dan MAPE.

3.6. Implementasi

1. Mempersiapkan Proses Prediksi
Setelah model SARIMA berhasil dimuat dan tanggal terakhir dari data training diketahui, proses dilanjutkan dengan melakukan prediksi. Pada bagian ini, model digunakan untuk meramalkan penjualan selama 7 bulan ke depan dengan perintah `model_loaded.forecast(steps=7)`. Hasilnya disimpan dalam variabel `forecast_7`.
2. Visualisasi Hasil Prediksi



Gambar 27 Visualisasi Hasil Prediksi
Sumber : Penelitian Mandiri 2025

Grafik di atas menampilkan perbandingan antara data historis penjualan 12 bulan terakhir dan hasil prediksi model SARIMA untuk 7 bulan ke depan.

Data Historis (2024-05 hingga 2025-04)

- Direpresentasikan oleh garis biru.
- Menunjukkan pola penjualan yang fluktuatif namun berulang secara musiman, dengan puncak tertinggi sekitar pertengahan tahun 2024 (bulan Juli–Agustus), dan penurunan signifikan pada akhir 2024.
- Pola ini mengindikasikan adanya siklus musiman tahunan dalam data penjualan.

Hasil Prediksi (2025-05 hingga 2025-11)

- Direpresentasikan oleh garis oranye.
- Model SARIMA memproyeksikan peningkatan penjualan dimulai dari bulan Mei 2025, dengan puncak tertinggi diperkirakan terjadi pada bulan Juli dan Agustus 2025.

- Setelah mencapai puncak, model memprediksi penurunan bertahap hingga bulan November 2025.

4. KESIMPULAN

Berdasarkan hasil penelitian dan analisis data penjualan pipa HDPE di PT. Awi Jaya Indotama, maka dapat disimpulkan beberapa hal sebagai berikut:

1. Penelitian ini berhasil menerapkan model SARIMA(0,1,0)(1,0,0,12) dalam melakukan prediksi penjualan bulanan pipa HDPE secara time series. Pemilihan model ini berdasarkan nilai AIC terendah.
2. Berdasarkan analisis ACF dan PACF, serta identifikasi parameter menggunakan pendekatan CRISP-DM, diperoleh model SARIMA dengan parameter terbaik. Model ini menunjukkan performa yang cukup baik dalam melakukan prediksi terhadap data uji, dengan nilai evaluasi yang tergolong akurat.
3. Aplikasi prediksi berbasis web menggunakan Flask berhasil dibangun untuk memvisualisasikan hasil prediksi penjualan secara interaktif untuk melihat tren historis, proyeksi penjualan.

DAFTAR PUSTAKA

- R. A. Pramono et al., "Microplastic Pollution from Polyethylene Bags in Coastal Areas of Indonesia," *Marine Pollution Bulletin*, vol. 186, p. 114456, 2023. Dan Manajemen STMIK, (November), 1–5.
- T. H. Noor, A. M. Almars, M. Alwateer, M. Almaliki, I. Gad, and E.-S. Atlam, "SARIMA: A Seasonal Autoregressive Integrated Moving Average Model for Crime Analysis in Saudi Arabia," *Electronics*, vol. 11, no. 23, art. 3986, Nov. 2022, doi: 10.3390/electronics11233986.