

A photograph of a professional meeting. In the foreground, a man in a dark suit and striped tie is gesturing with his hands. Behind him, a woman with her hair pulled back is looking down at something. To the right, another person's face is partially visible, smiling. Overlaid on the top right is a white grid with a blue bar chart showing increasing values from left to right, and a red line graph showing an upward trend.

Project 1 - SALES FORECASTING

by 360

Meet the Team



Yanto



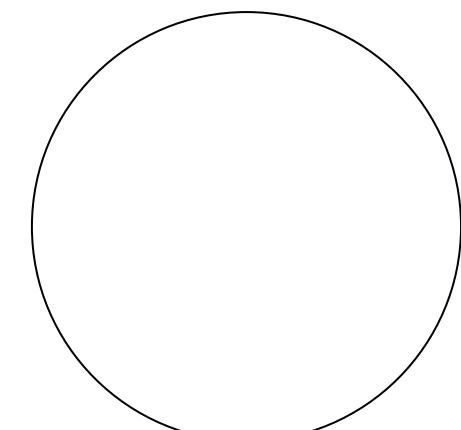
Muhammad Sutisna



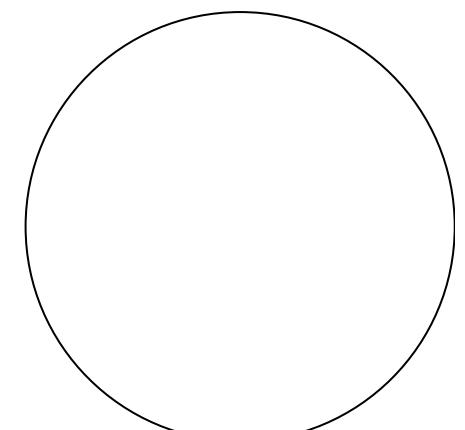
Vika Oktarina



Ryandito Mahendradani



A. Mambaus Sholihin



Mukhammad Fatkhur Rozi



Background & Problem Statement

Dalam usaha toko groceries persediaan stock barang menjadi masalah tersendiri bagi pelaku usaha. Sering kali toko akan mengalami kerugian apabila toko tidak dapat menyiapkan jumlah barang sesuai dengan permintaan. Alhasil banyak barang tidak terjual dan menjadi kadaluarsa sehingga menimbulkan kerugian, atau sebaliknya barang yang laku dijual tapi toko malah kehabisan stock sehingga tidak dapat meraup keuntungan.

Oleh karena itu, penyediaan stock yang pas dalam waktu yang tepat sangat diperlukan untuk memaksimalkan potensi laba usaha.

Objectives & Scope

Objectives

Sesuai dengan latar belakang permasalahan yang ada. Maka untuk itu Tim 360 mengembangkan sistem AI untuk meramalkan penjualan toko, sehingga pemilik toko dapat menyiapkan stock sesuai dengan permintaan dan memaksimalkan laba

Scope

Peramalan dibatasi dalam scope sebagai berikut:

1. Data yang dianalisa kategori Groceries I pada toko groceries kode 5
2. Peramalan penjualan dengan sistem AI model time series ARIMA & LSTM



Data Collection & Preparation

Dalam melakukan eksplorasi data, ditemukan hal-hal berikut:



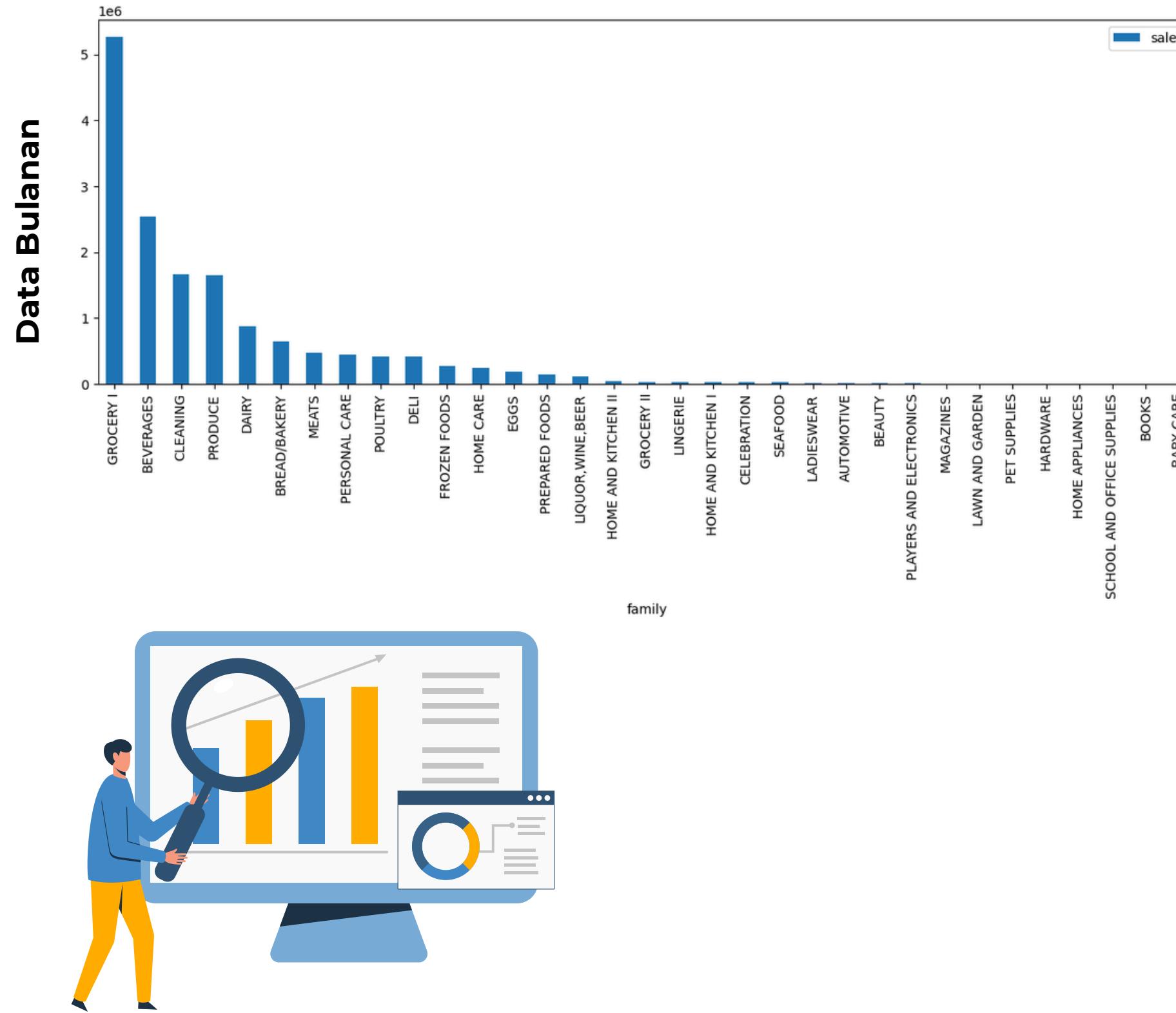
Terdapat 5 fitur dalam data yaitu:

1. Family: Jenis barang
2. Store_nbr: Nomor cabang
3. Sales: Jumlah penjualan
4. Onpromotion: Data apakah barang dalam promosi atau tidak
5. Dcoilwtico: Harga minyak mentah

Persiapan data dengan sebagai berikut:

1. Menghapus kolom store_nbr karena hanya kode toko dan memang dataset sudah terfilter ke kolom 5 sesuai objectives dan scope
2. Menghapus kolom Dcoilwtico karena banyak data kosong
3. Memilih kategori produk / Family yang paling layak diteliti
4. Melakukan uji korelasi antara OnPromotion dengan Sales

Data Collection & Preparation



Data yang dipakai difokuskan pada kategori Groceries I dikarenakan:

1. Groceries (bahan pokok) merupakan kategori produk dengan permintaan yang relatif stabil dan frekuensi penjualan tinggi
2. Karena frekuensi penjualan yang tinggi, kategori ini memberikan kontribusi pendapatan yang tinggi juga
3. Dengan volume penjualan yang tinggi, groceries berisiko tinggi terhadap overstock atau stockout
4. Kategori groceries sering kali memiliki data historis yang lebih lengkap karena konsisten dijual setiap hari. Ini sangat berguna untuk time series forecasting menggunakan model seperti LSTM, ARIMA, atau SARIMA.
5. Pola konsumsi groceries dapat merefleksikan tren ekonomi makro atau perubahan perilaku konsumen.

Data Collection & Preparation

Dalam hal uji korelasi variabel OnPromotion dengan Penjualan Groceries I, tidak ada penemuan korelasi yang terlalu signifikan dengan nilai korelasi tidak 0.44 tidak lebih dari 0.5

| | sales | onpromotion |
|--------------------|--------------|--------------------|
| sales | 1.000000 | 0.448904 |
| onpromotion | 0.448904 | 1.000000 |



Hal ini mungkin disebabkan, kategori Groceries adalah kebutuhan pokok yang dimana ada atau tidak adanya promosi orang akan tetap butuh dan membeli.

Model Development

Model Yang Digunakan



1

ARIMA (AutoRegressive Integrated Moving Average):

Cocok untuk pola linier dengan asumsi stasioneritas.



2

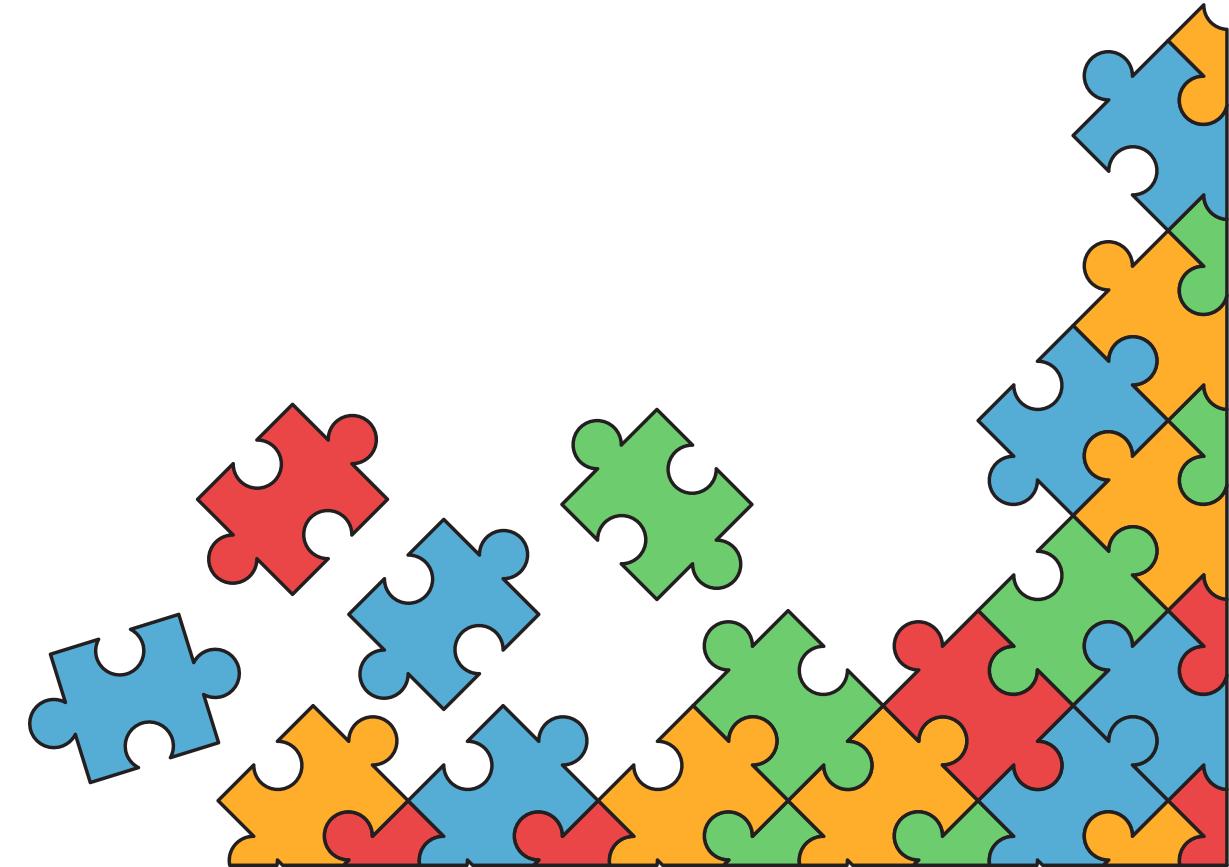
LSTM (Long Short-Term Memory):

Recurrent Neural Network yang unggul untuk pola time series kompleks dengan memori jangka panjang.

Model Development

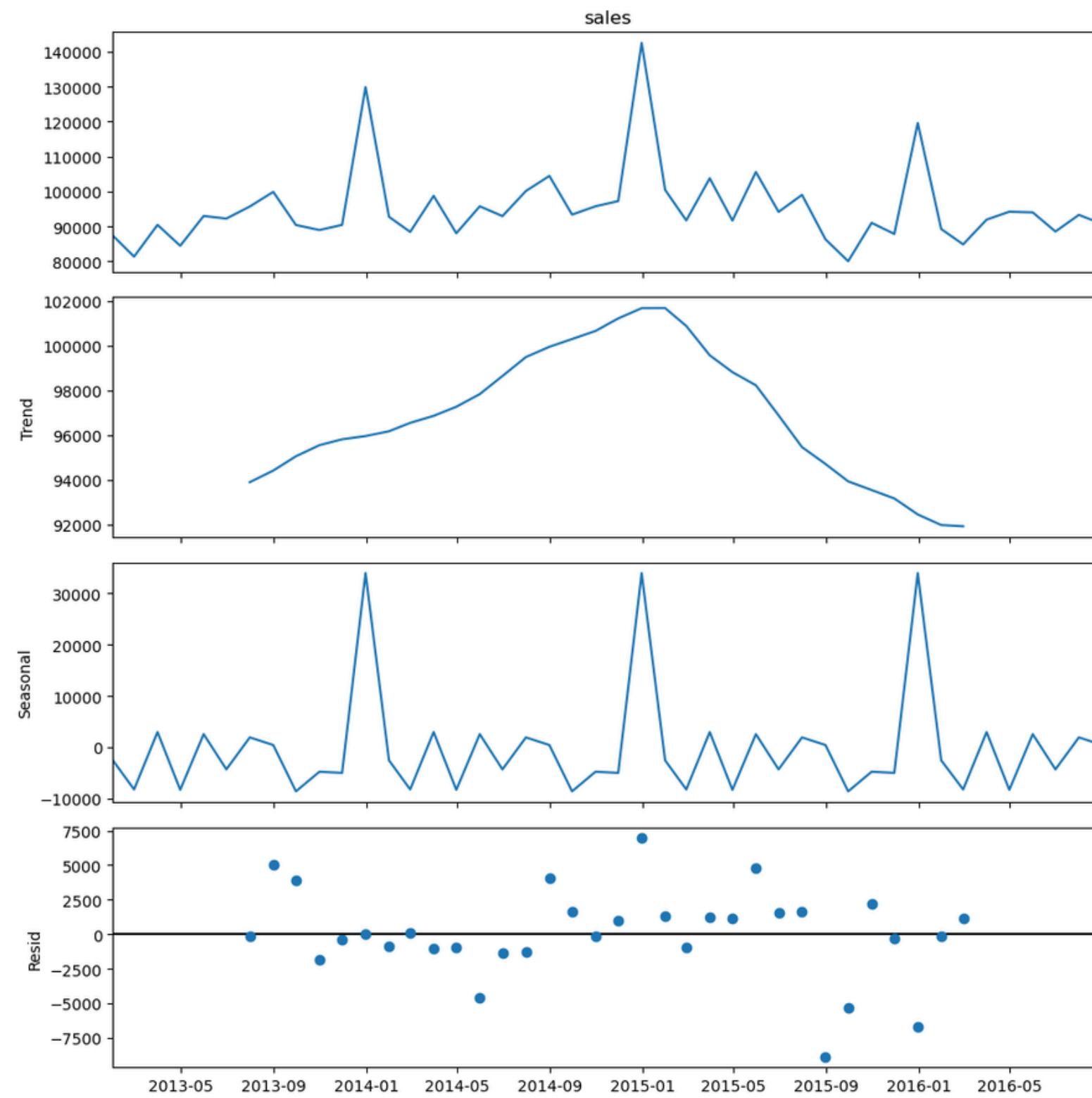
ARIMA MODEL (dengan Data Bulanan & Harian)

- Plot data time series
- Uji stasioneritas
- Diferensiasi Data
- Analisis ACF & PACF
- Bangun model ARIMA
- Training Model ARIMA (p,d,q)
- Forecasting
- Evaluasi Model

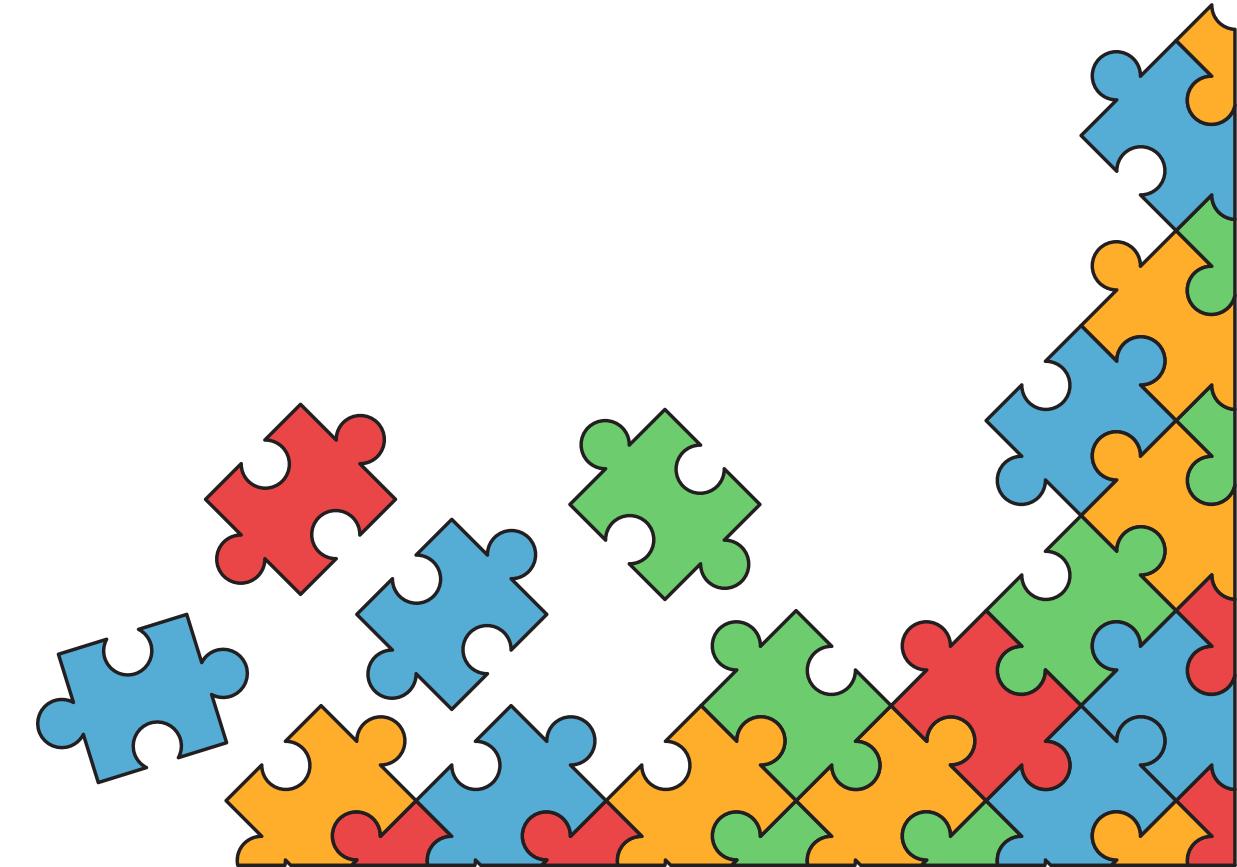


Model Development

ARIMA MODEL (dengan Data Bulanan)



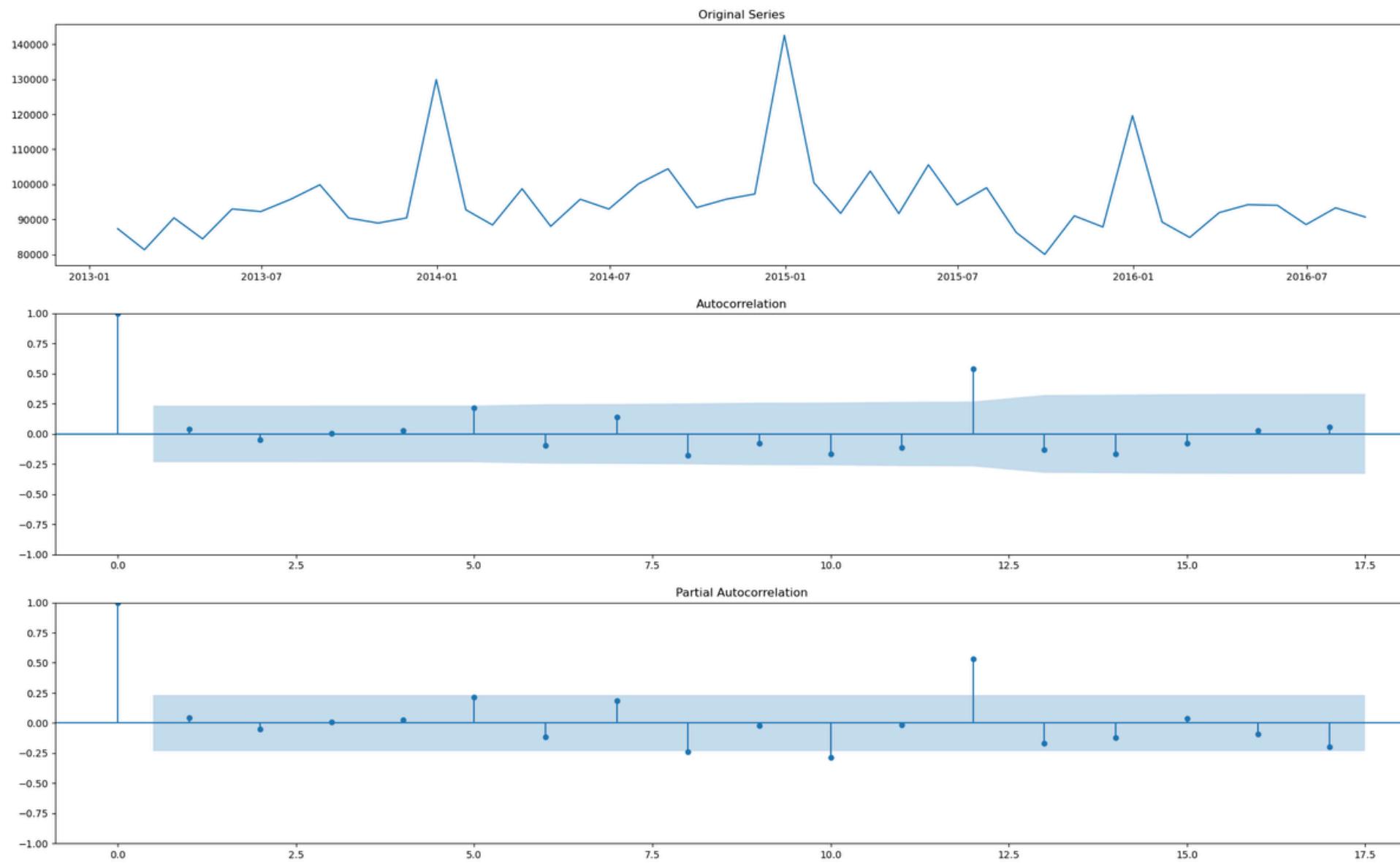
Berdasarkan diagram disamping menunjukkan bahwa data tersebut memiliki seasonal, maka akan digunakan metode SARIMA



Model Development

ARIMA MODEL (dengan Data Bulanan)

MODEL ARIMA (p,d,q) (P,D,Q,s)



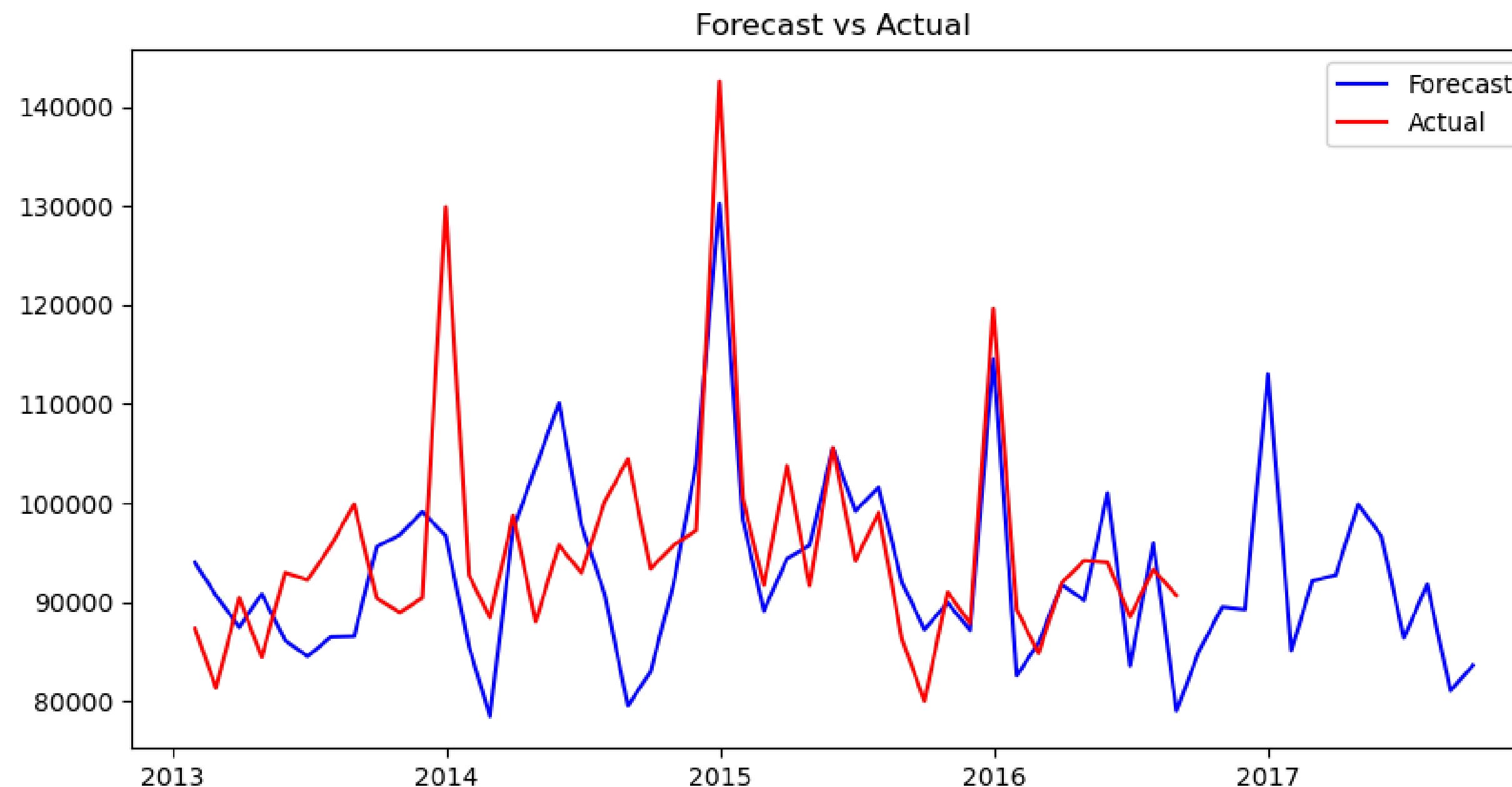
Kesimpulan dari grafik ACF dan PACF :

1. Data awal atau data asli sudah stationer, karena p-value < 0.05
2. Data hasil pengecekan grafik Decompose menunjukkan Seasonal dengan $s = 12$
3. Data bisa dikatakan juga stationer bisa dilihat dari hasil differencing p-value < 0.05,
4. Model SARIMA (p,d,q) (P, D, Q, S) dapat dipilih dengan data :
 1. p, bisa diambil dari PACF, misal $p = 8$
 2. q, dari ACF bisa diambil $q = 5$
 3. d, karena tidak dilakukan differencing maka $d = 0$
 4. P, index diambil dari PACF ke 1, maka $P = 1$
 5. D, karena tidak dilakukan differencing maka $D = 0$
 6. Q, index diambil dari ACF ke 1, maka $Q = 1$
 7. S, diambil dari seasonal tiap 12 bulan, maka $s = 12$
5. kesimpulan SARIMA (8,0,5)(1,0,1,12)

Model Development

ARIMA MODEL (dengan Data Bulanan)

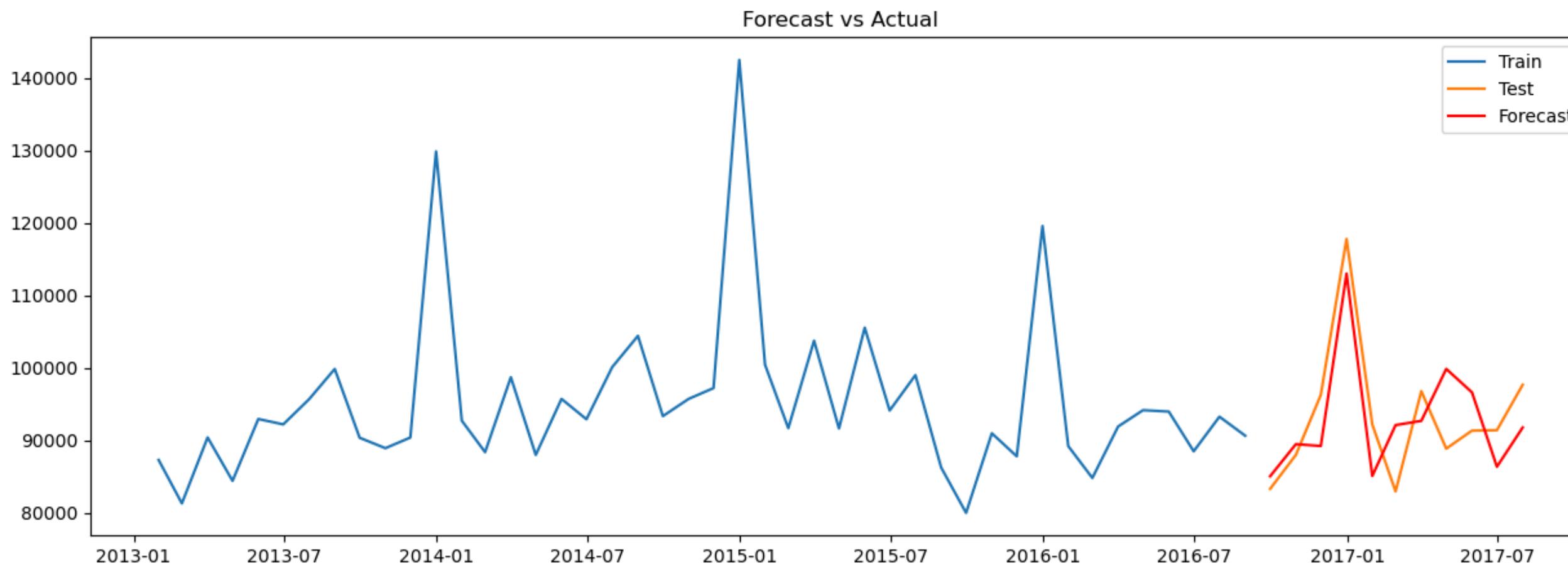
Hasil Metode ARIMA (Data Bulanan)



Model Development

ARIMA MODEL (dengan Data Bulanan)

Hasil Metode ARIMA
(Data Bulanan)

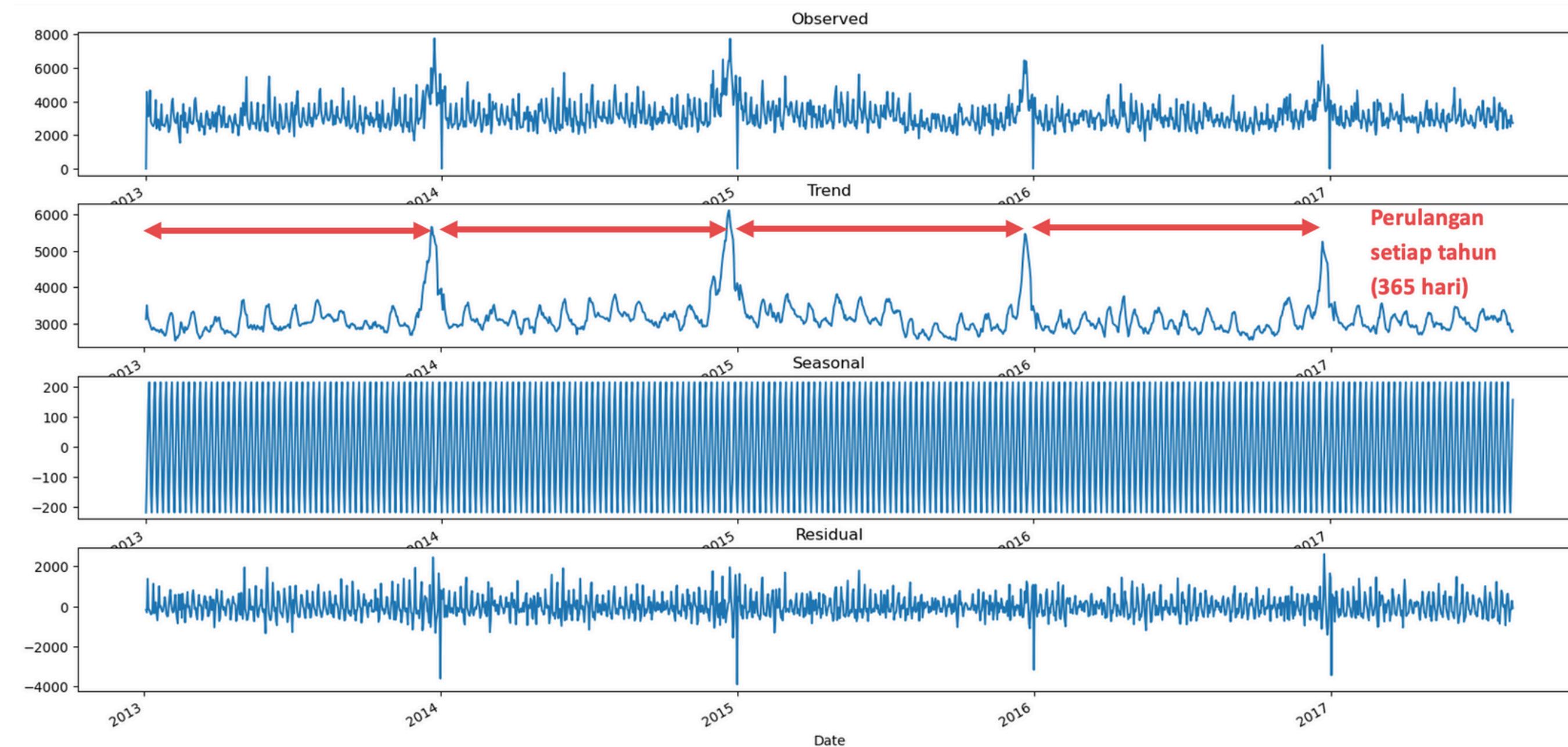


Berdasarkan hasil nilai MAPE : 6% < 10%, maka bisa dianggap cukup baik.

Model Development

ARIMA MODEL (dengan Data Harian)

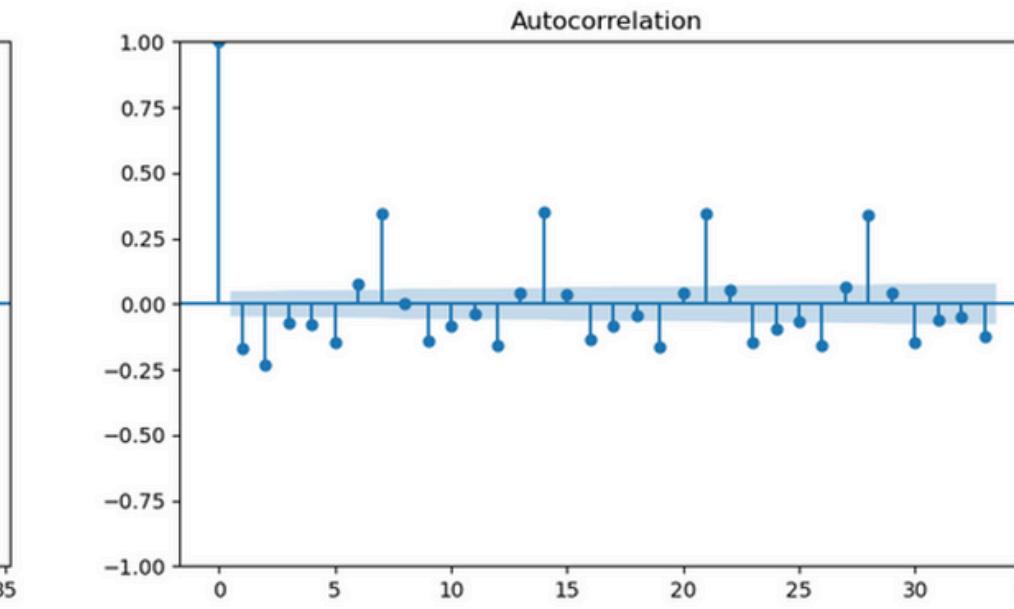
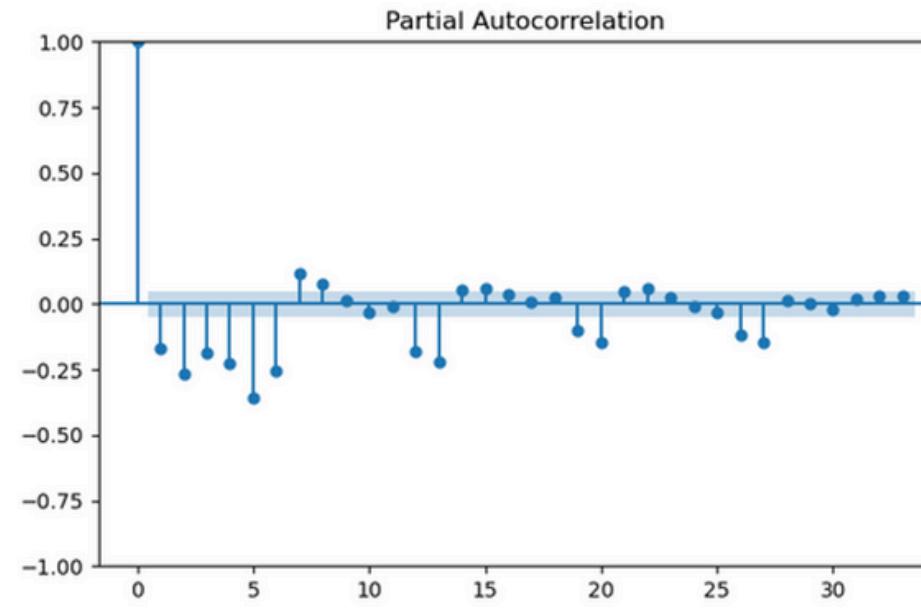
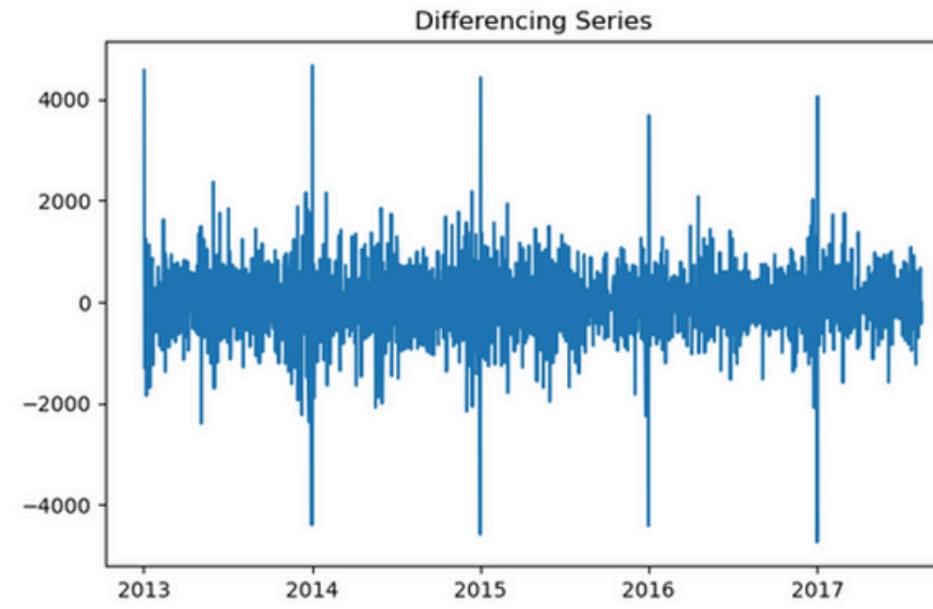
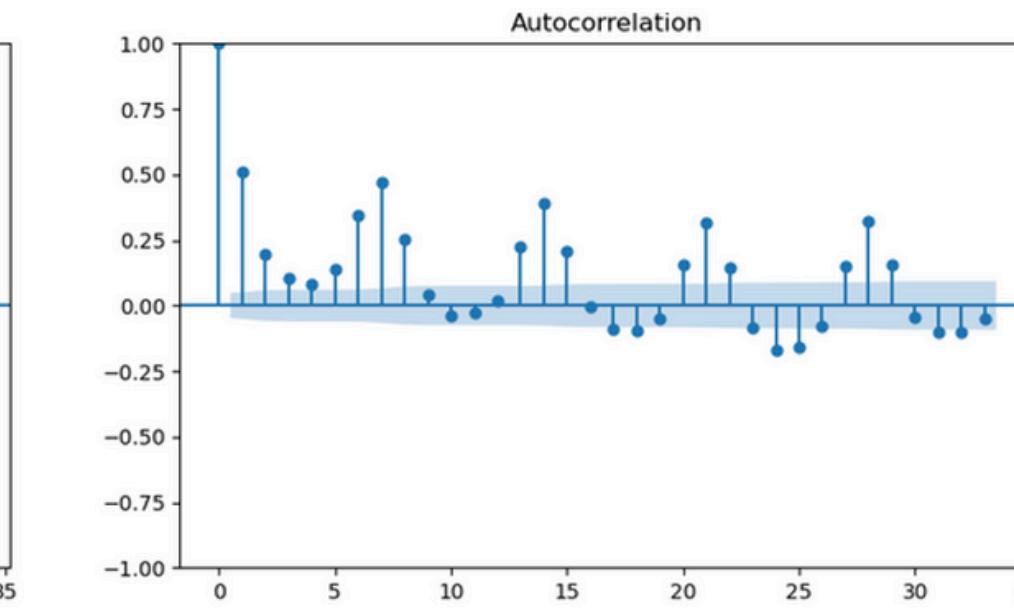
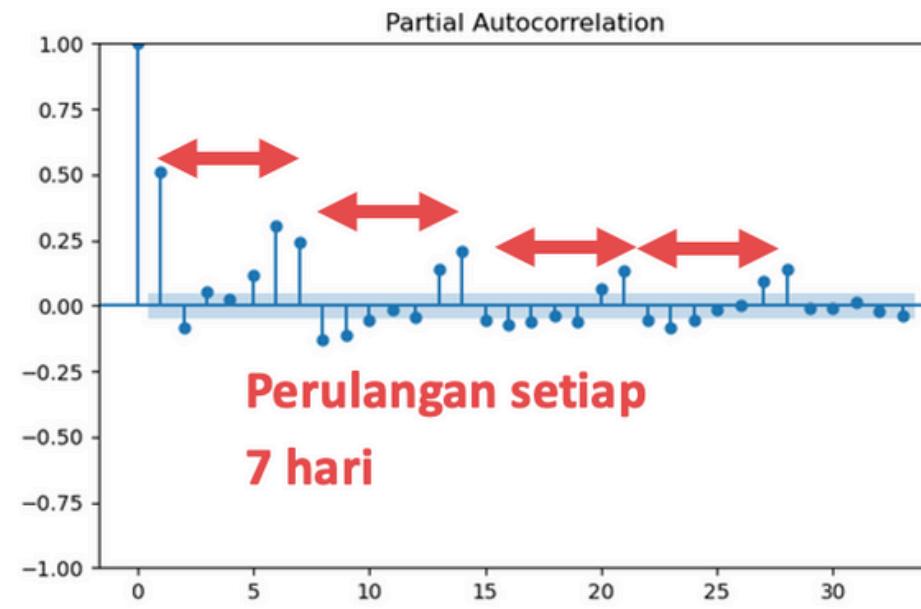
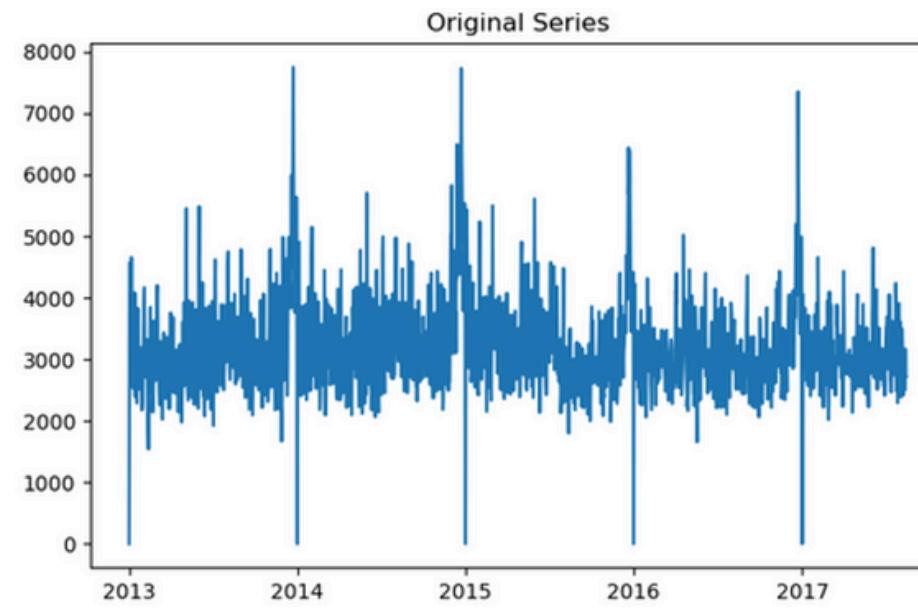
Eksplorasi Data (Data Harian)



Model Development

ARIMA MODEL (dengan Data Harian)

MODEL ARIMA (p,d,q) (P,D,Q,s)



Model Development

ARIMA MODEL (dengan Data Harian)

Dari PACF, kita bisa mengambilkan nilai AR(p) = 1,

Dari ACF, kita bisa mengambilkan nilai MA(q) = 1,

Dari hasil uji stationarity, kita bisa mengambilkan nilai d = 0,

Dari PACF, kita bisa mengambil kesimpulan seasonal = 7

Dari Data Decompose 'Trend', kita bisa mengambil kesimpulan seasonal = 365

Berdasarkan data tersebut, diputuskan bahwa akan dilakukan 3 metode ARIMA dalam pemodelan:

1.ARIMA Orde (1,0,1), Orde Seasonality (1,0,1,365)

2.ARIMA Orde (1,0,1), Orde Seasonality (1,0,1,7)

3.Auto ARIMA

Results

ARIMA MODEL (dengan Data Harian)

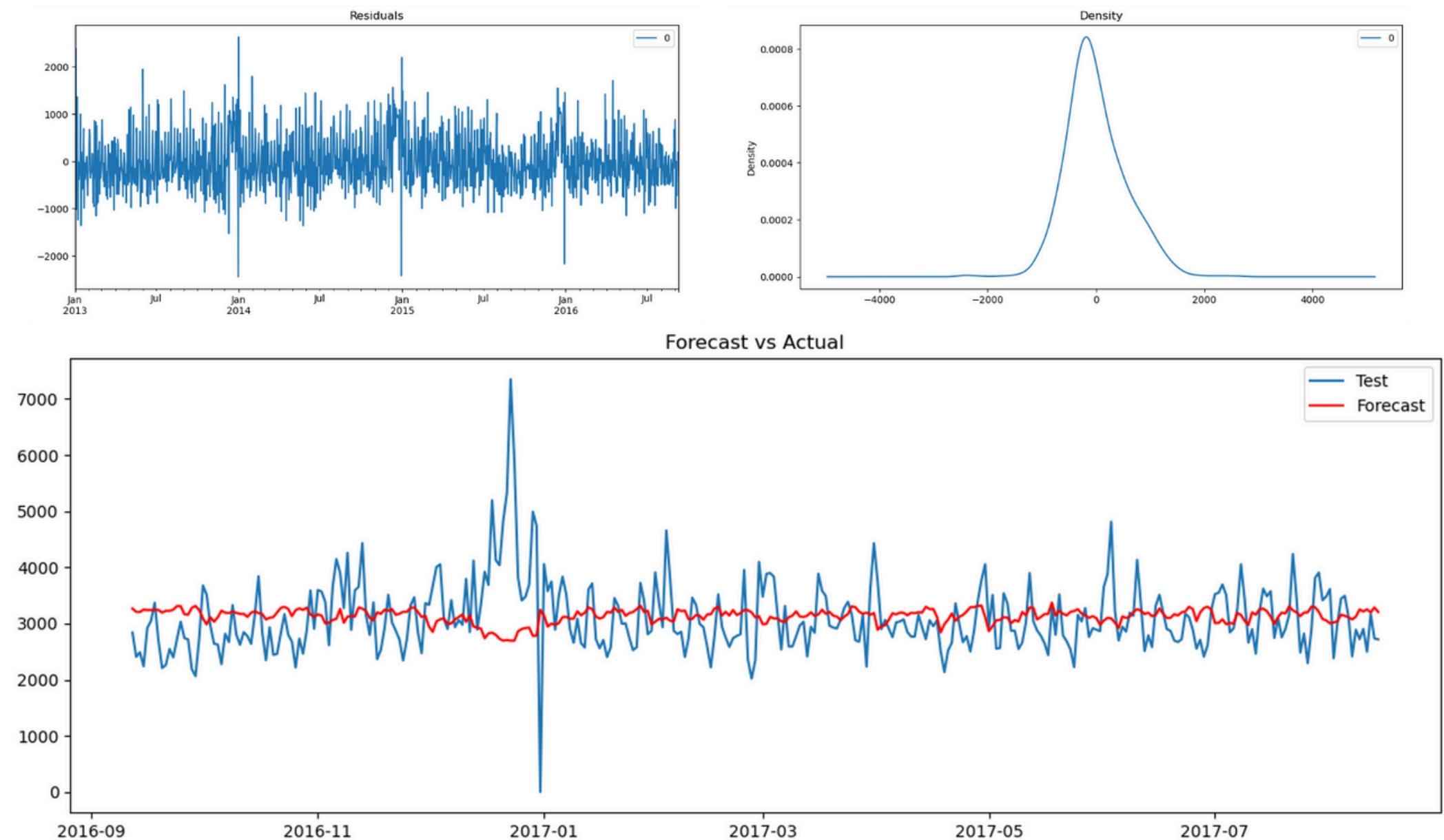
MODEL ARIMA - 365

```
SARIMAX Results
=====
Dep. Variable: Sales No. Observations: 1347
Model: ARIMA(1, 0, 1)x(1, 0, 1, 365) Log Likelihood: 10456.580
Date: Fri, 30 May 2025 AIC: 20872.611
Time: 18:24:01 BIC: 20903.845
Sample: 01-01-2013 HQIC: 20884.399
Covariance Type: opg
=====

coef std err z P>|z| [0.025 0.975]
-----
const 3133.8059 28.804 108.799 0.000 3077.352 3190.260
ar.L1 0.3509 0.041 8.466 0.000 0.270 0.432
ma.L1 0.2792 0.042 6.634 0.000 0.197 0.362
ar.S.L365 0.3355 0.281 1.195 0.232 -0.215 0.886
ma.S.L365 -0.4925 0.276 -1.785 0.074 -1.033 0.048
sigma2 3.106e+05 1.12e+04 27.720 0.000 2.89e+05 3.33e+05
=====

Ljung-Box (L1) (Q): 0.03 Jarque-Bera (JB):
Prob(Q): 0.87 Prob(JB): 110.57
Heteroskedasticity (H): 0.71 Skew: 0.00
Prob(H) (two-sided): 0.00 Kurtosis: 4.14
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```



Results

ARIMA MODEL (dengan Data Harian)

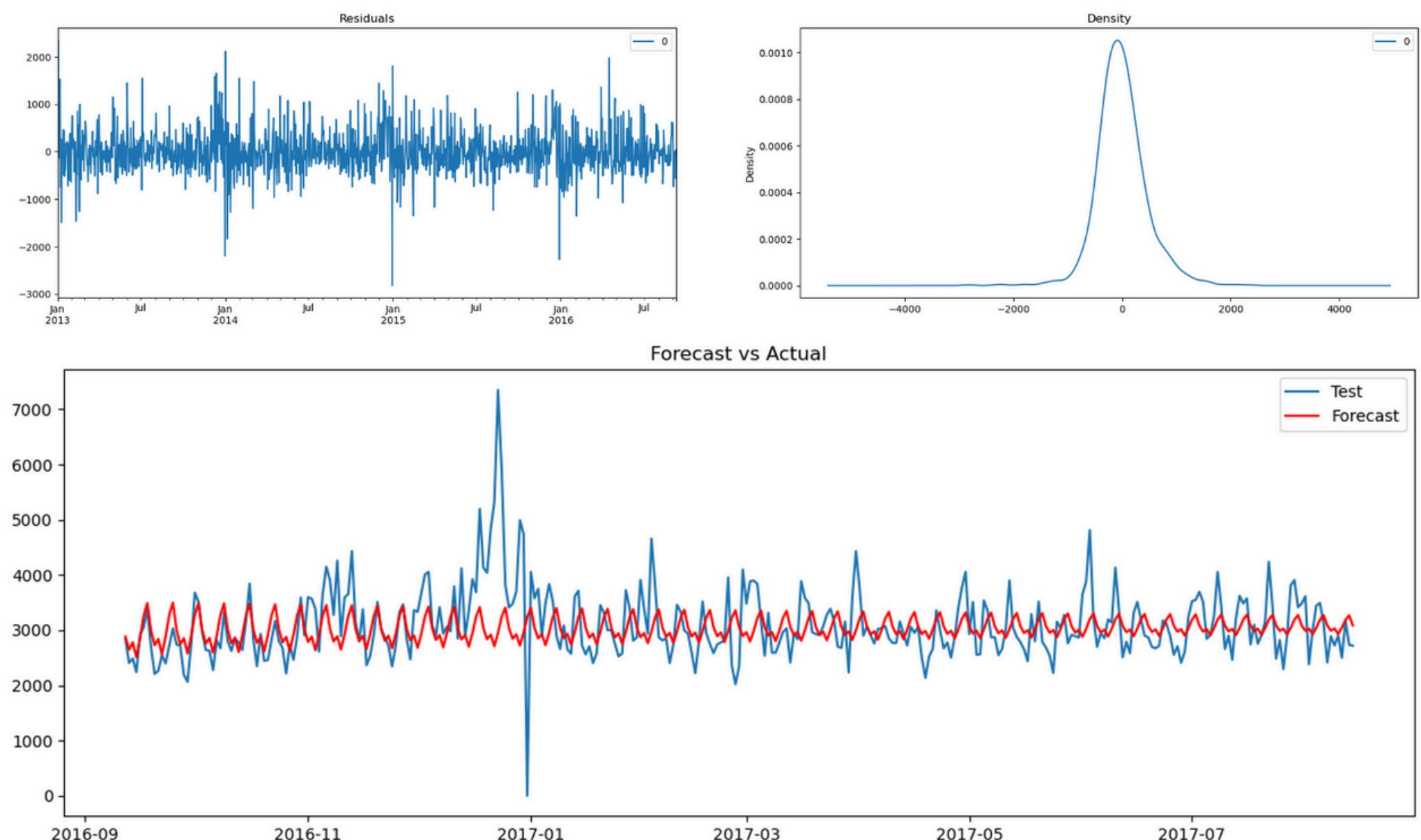
MODEL ARIMA - 7

```
SARIMAX Results
=====
Dep. Variable: Sales No. Observations: 1347
Model: ARIMA(1, 0, 1)x(1, 0, 1, 7) Log Likelihood: -10155.832
Date: Fri, 30 May 2025 AIC: 20323.664
Time: 18:24:06 BIC: 20354.897
Sample: 01-01-2013 HQIC: 20335.362
Covariance Type: opg
=====

coef std err z P>|z| [0.025 0.975]
-----
const 3133.7039 169.536 18.484 0.000 2801.419 3465.989
ar.L1 0.7240 0.031 23.250 0.000 0.663 0.785
ma.L1 -0.3281 0.040 -8.305 0.000 -0.406 -0.251
ar.S.L7 0.9781 0.005 192.401 0.000 0.968 0.988
ma.S.L7 -0.8028 0.018 -44.962 0.000 -0.838 -0.768
sigma2 2.059e+05 5264.906 39.116 0.000 1.96e+05 2.16e+05
=====

Ljung-Box (L1) (Q): 0.74 Jarque-Bera (JB): 650.53
Prob(Q): 0.39 Prob(JB): 0.00
Heteroskedasticity (H): 0.84 Skew: 0.21
Prob(H) (two-sided): 0.06 Kurtosis: 6.38
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```



Results

ARIMA MODEL (dengan Data Harian)

MODEL ARIMA - Auto

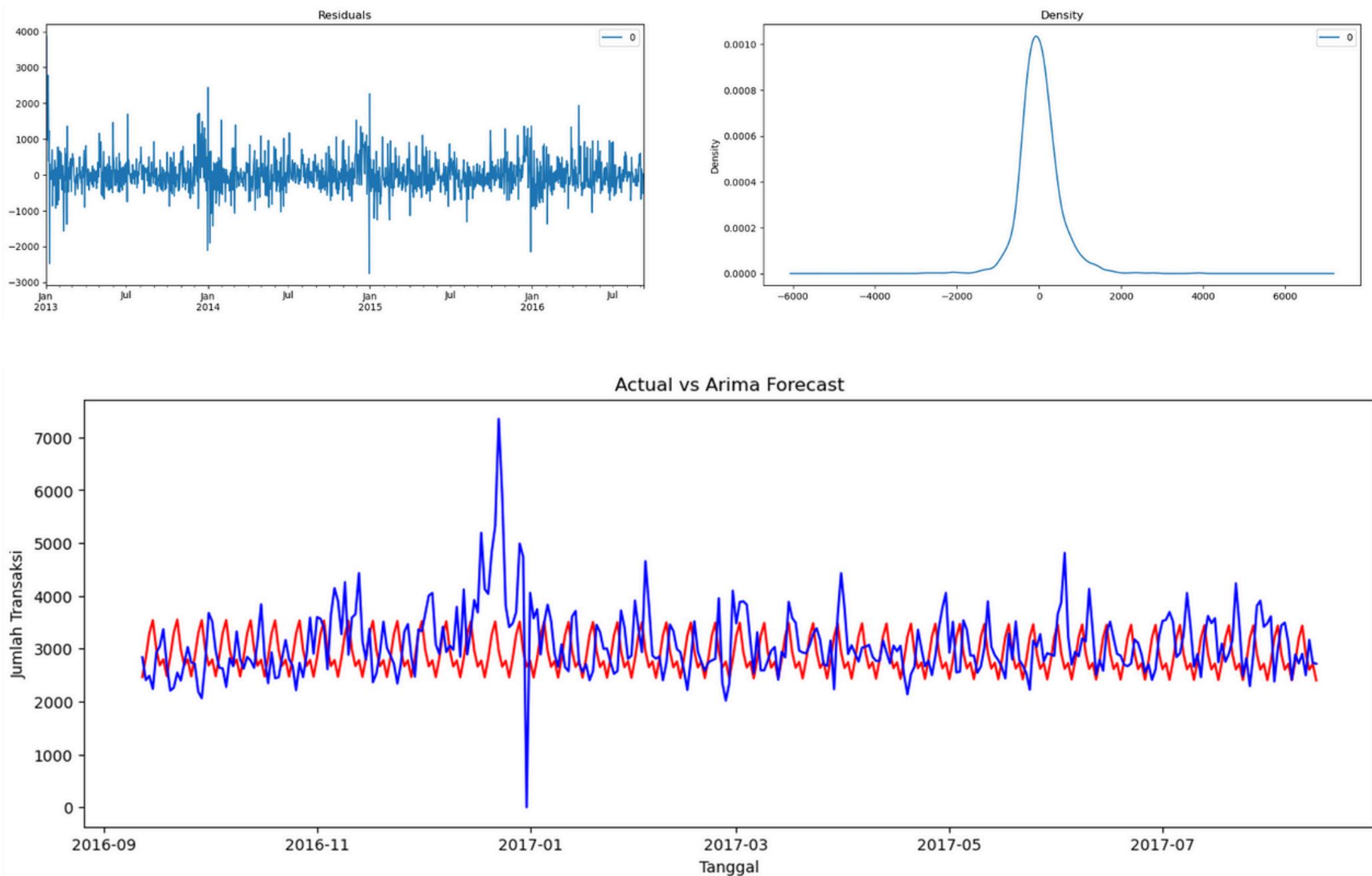
```
Best model: ARIMA(1,0,0)(2,0,2)[7]
Total fit time: 149.300 seconds

SARIMAX Results
=====
Dep. Variable: y No. Observations: 1347
Model: SARIMAX(1, 0, 0)x(2, 0, [1, 2], 7) Log Likelihood: -10183.46
Date: Fri, 30 May 2025 AIC: 20378.92
Time: 18:26:36 BIC: 20410.16
Sample: 01-01-2013 HQIC: 20380.62
Covariance Type: opg
=====

coef std err z P>|z| [0.025 0.975]
-----
ar.L1 0.4981 0.018 27.483 0.000 0.463 0.534
ar.S.L7 0.4969 1.430 0.347 0.728 -2.307 3.300
ar.S.L14 0.5021 1.429 0.351 0.725 -2.299 3.304
ma.S.L7 -0.3164 1.424 -0.222 0.824 -3.108 2.476
ma.S.L14 -0.4274 1.174 -0.364 0.716 -2.729 1.874
sigma2 2.132e+05 5372.583 39.682 0.000 2.03e+05 2.24e+05
=====

Ljung-Box (L1) (Q): 4.01 Jarque-Bera (JB): 780.77
Prob(Q): 0.05 Prob(JB): 0.00
Heteroskedasticity (H): 0.82 Skew: 0.22
Prob(H) (two-sided): 0.03 Kurtosis: 6.70
=====

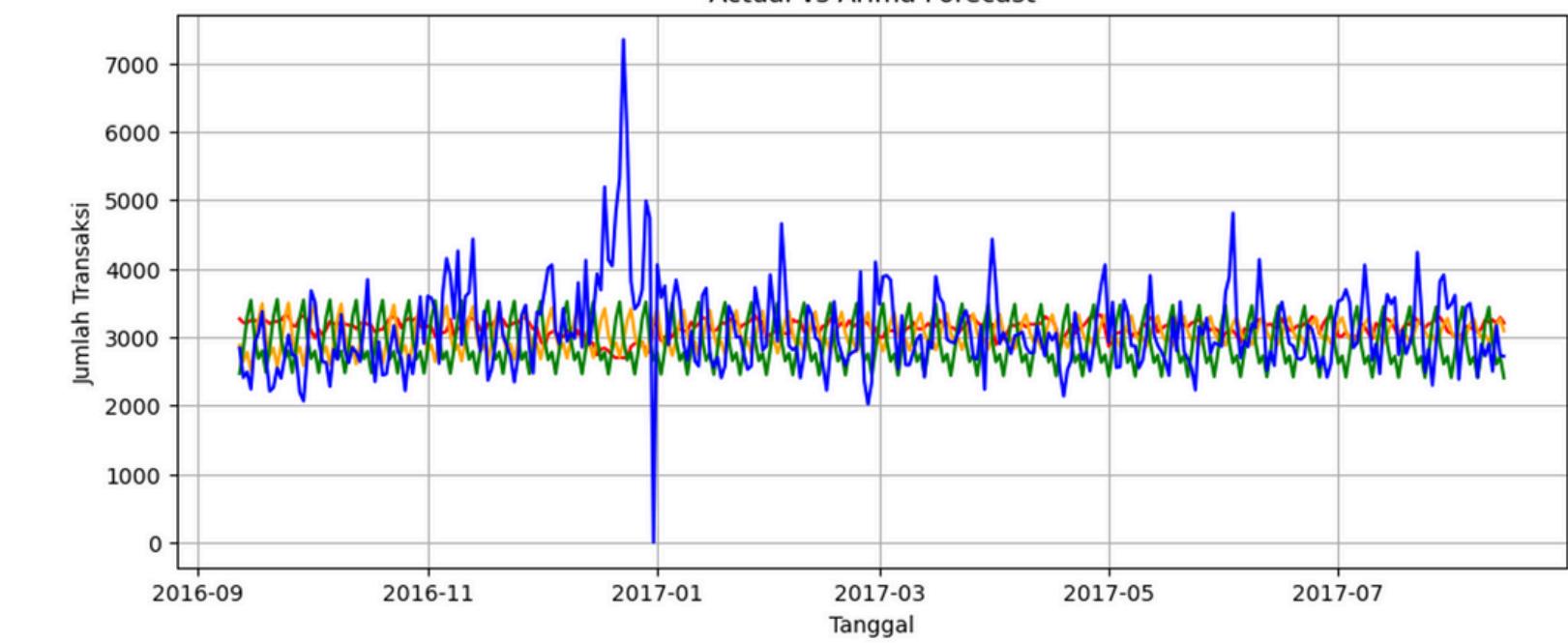
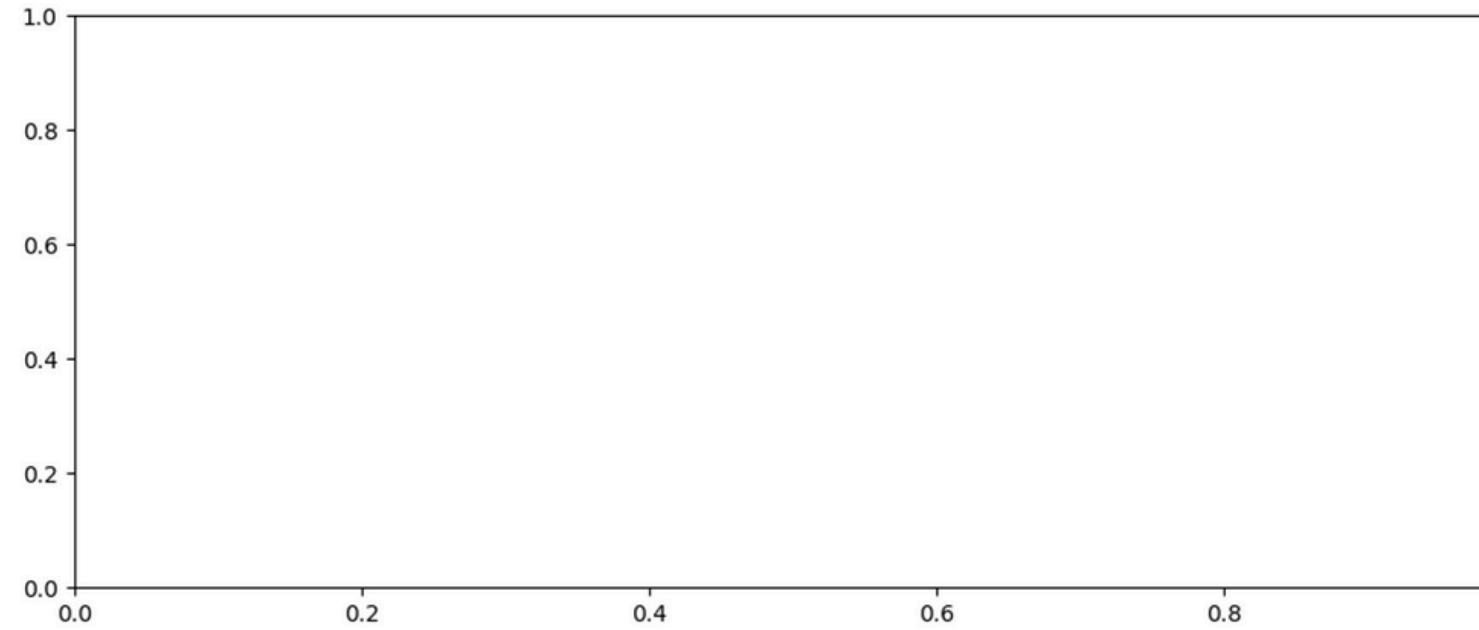
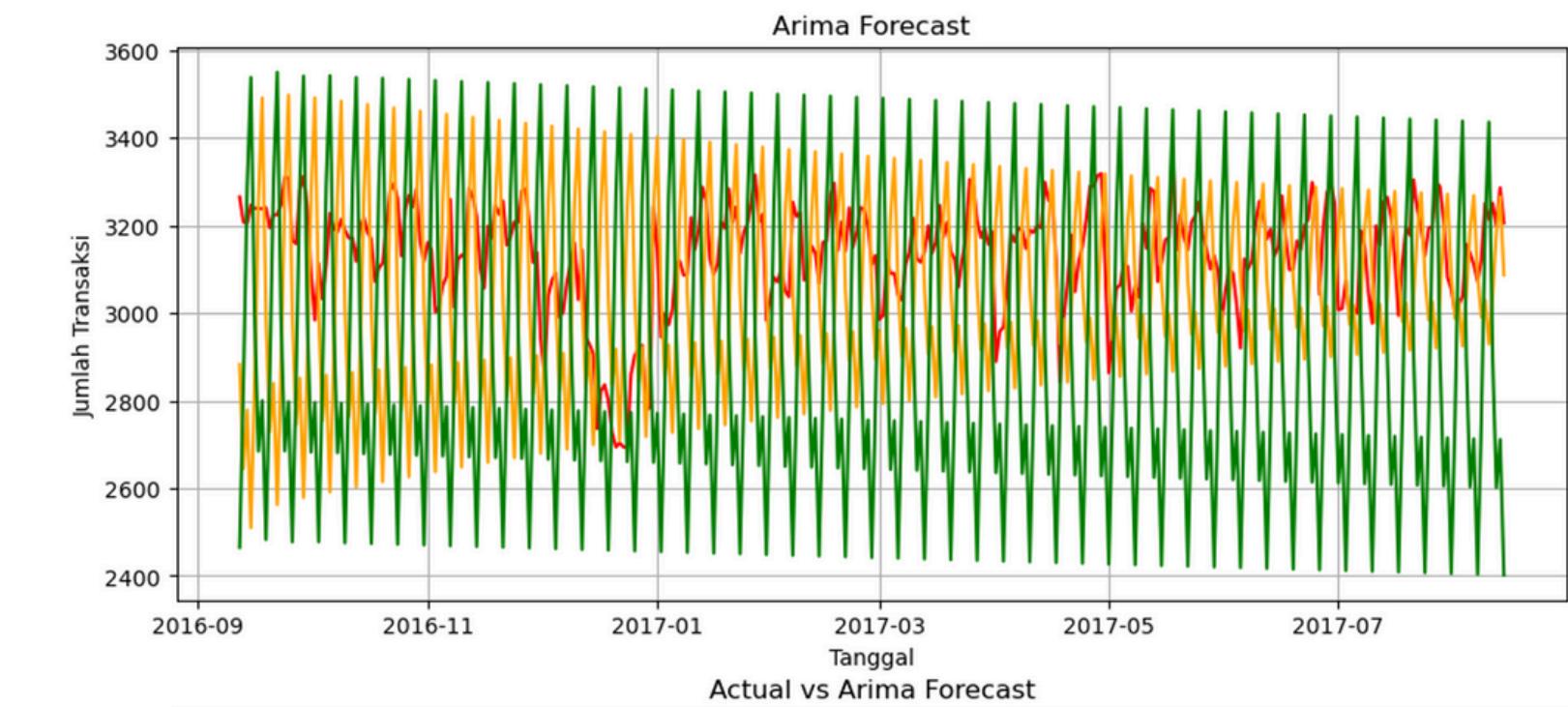
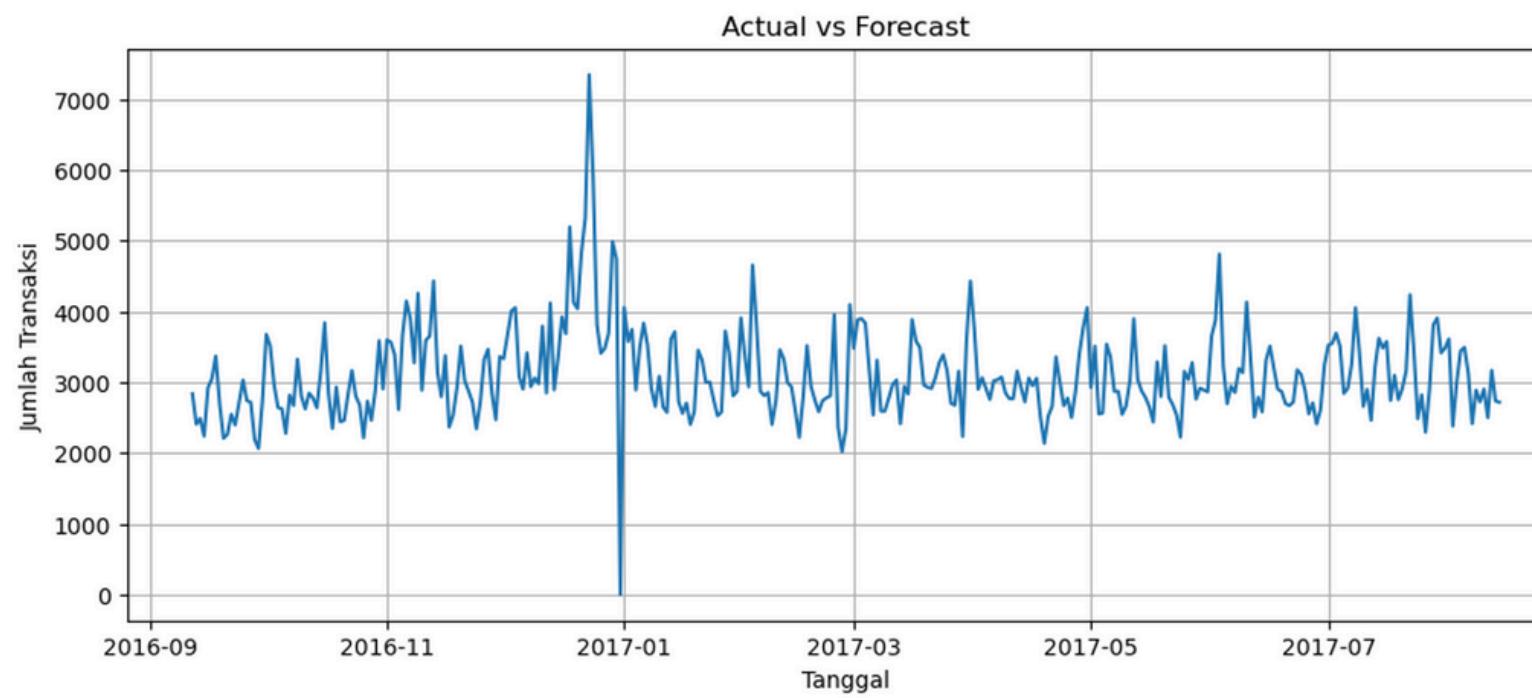
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```



Results

ARIMA MODEL (dengan Data Harian)

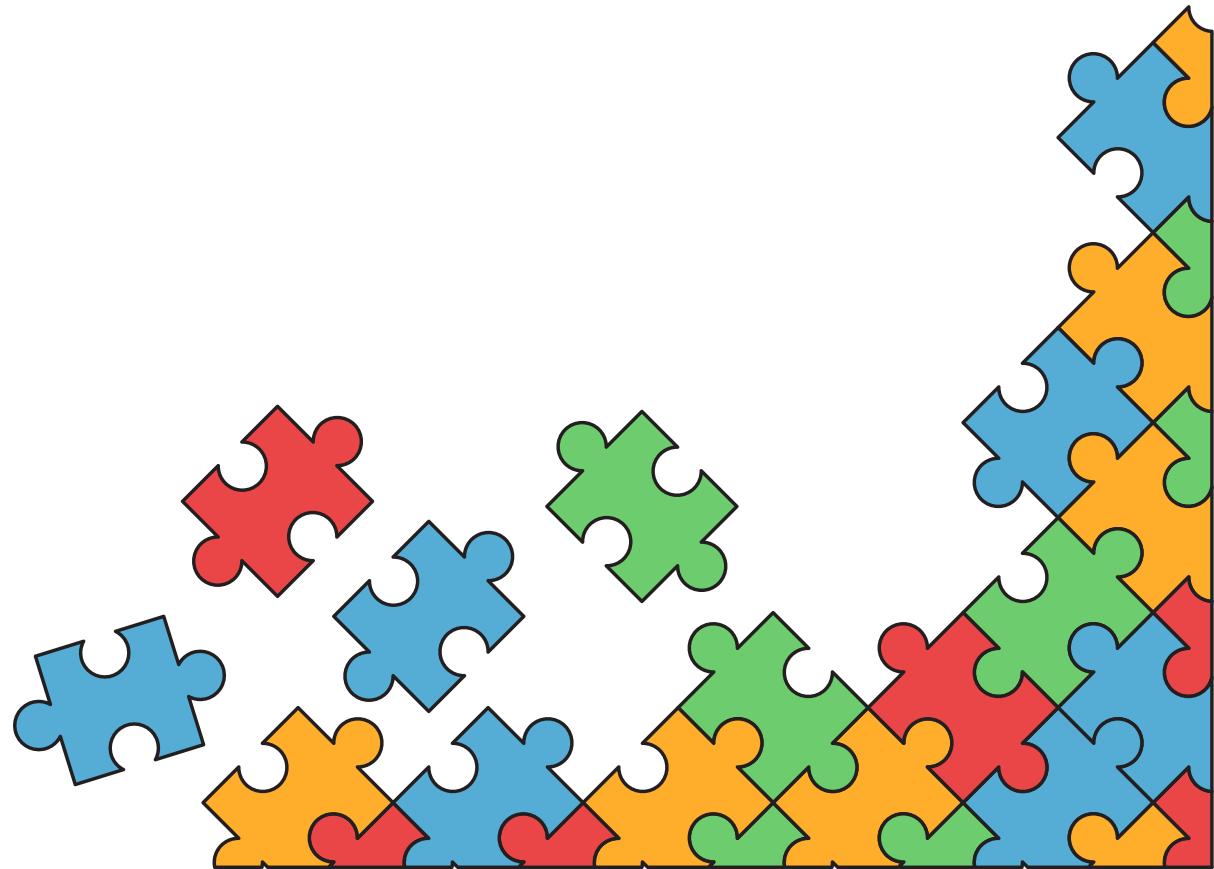
MODEL ARIMA - Comparison



Model Development

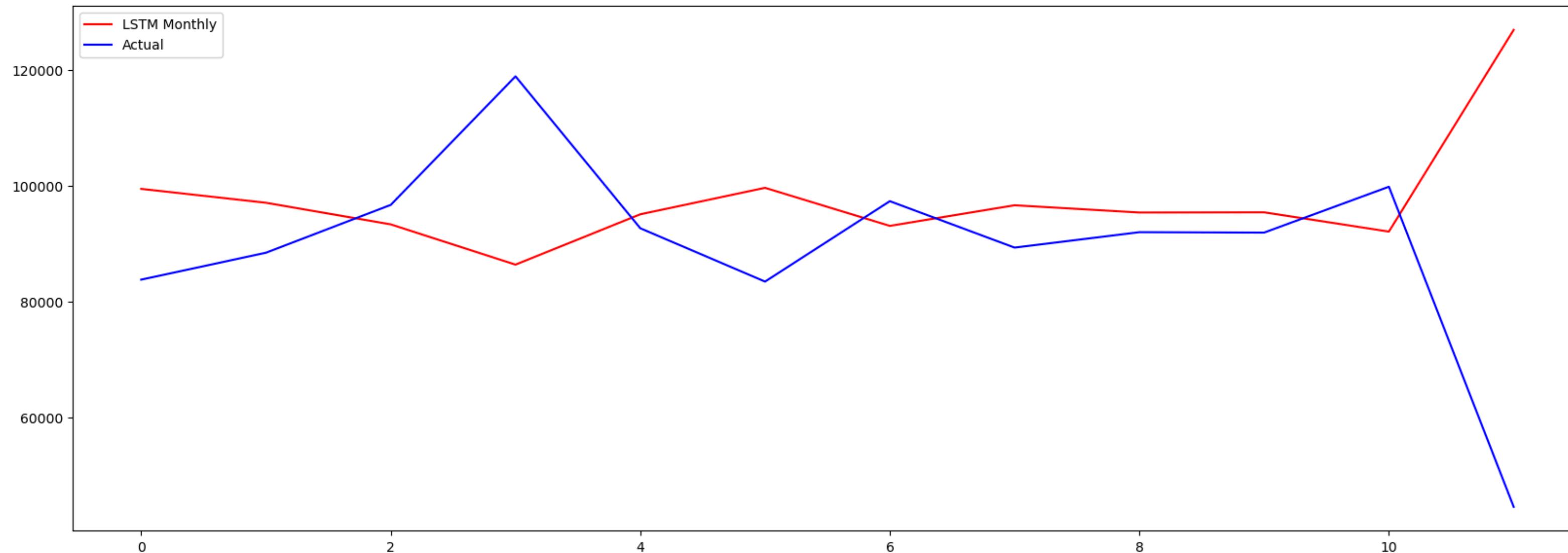
LSTM Model (dengan Data Bulanan & Harian)

- Preprocessing Data
- Reshape Input untuk LSTM
- Bangun Arsitektur Model LSTM
- Training Model
- Forecasting
- Evaluasi Model



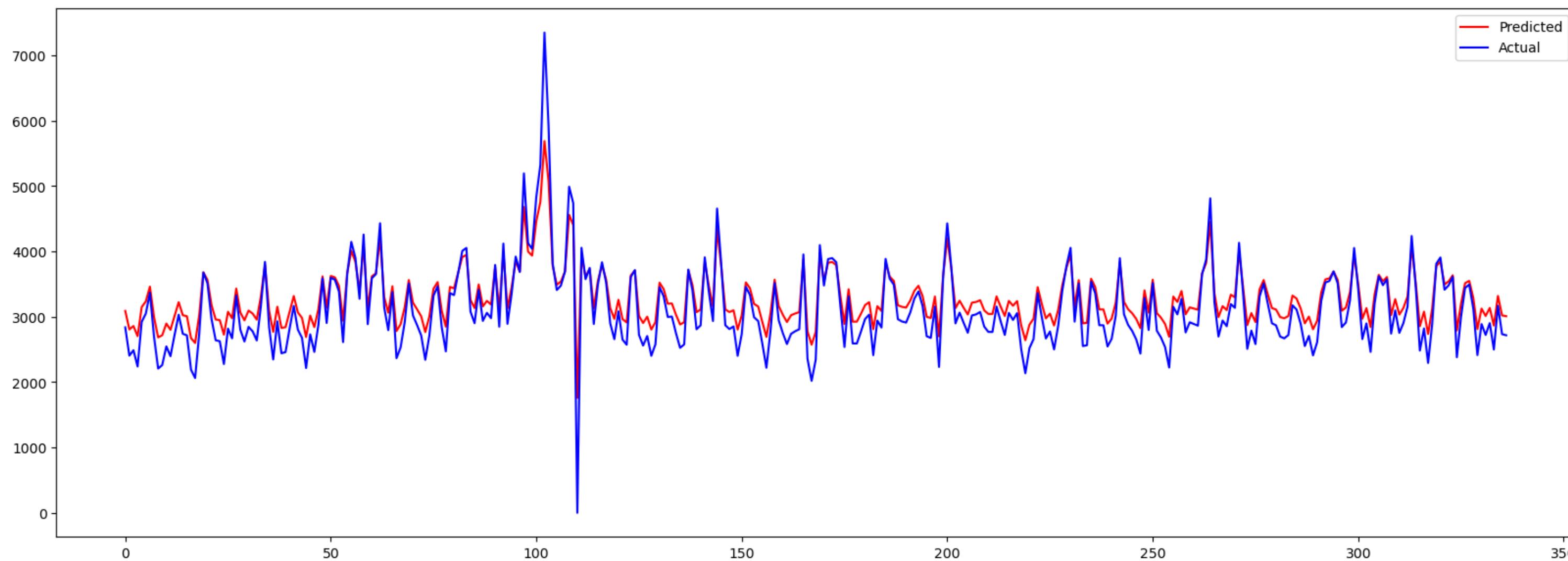
Model Development

LSTM Model (dengan Data Bulanan)



Model Development

LSTM Model (dengan Data Harian)



Results

ARIMA & LSTM Model Comparison

Comparison MAE , MAPE & RSME

MAPE ARIMA 365 = 15.99%

MAE ARIMA 365 = 509.51

RMSE ARIMA 365 = 701.40

MAPE ARIMA 7 = 13.02%

MAE ARIMA 7 = 434.05

RMSE ARIMA 7 = 635.69

MAPE Auto ARIMA = 18.09%

MAE Auto ARIMA = 589.30

RMSE Auto ARIMA = 774.28

MAPE LSTM = 7.80%

MAE LSTM = 226.39

RMSE LSTM = 285.35

MAPE ARIMA Monthly 12 = 19.12%

MAE ARIMA Monthly 12 = 12554.12

RMSE ARIMA Monthly 12 = 18900.42

MAPE ARIMA Monthly 7 = 17.38%

MAE ARIMA Monthly 7 = 11444.17

RMSE ARIMA Monthly 7 = 18345.31

MAPE Auto ARIMA Monthly = 14.98%

....

MAPE LSTM Monthly = 24.50%

MAE LSTM Monthly = 15621.96

RMSE LSTM Monthly = 26776.41

Results

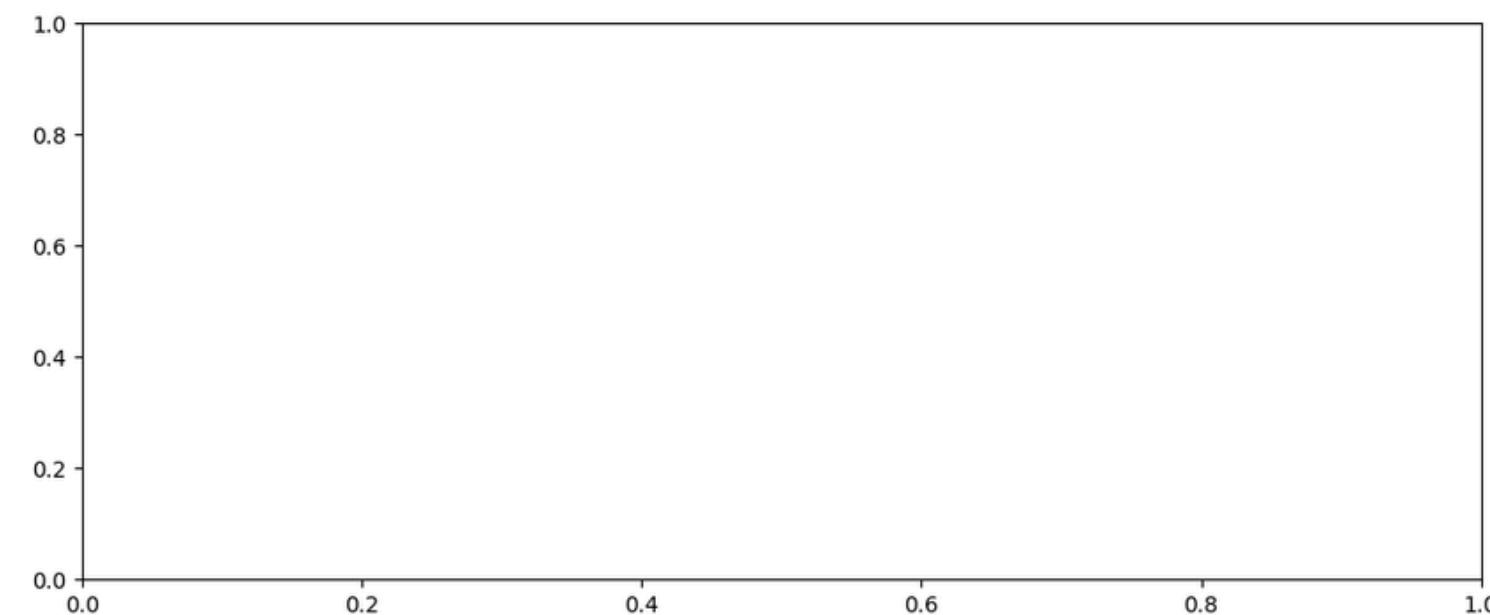
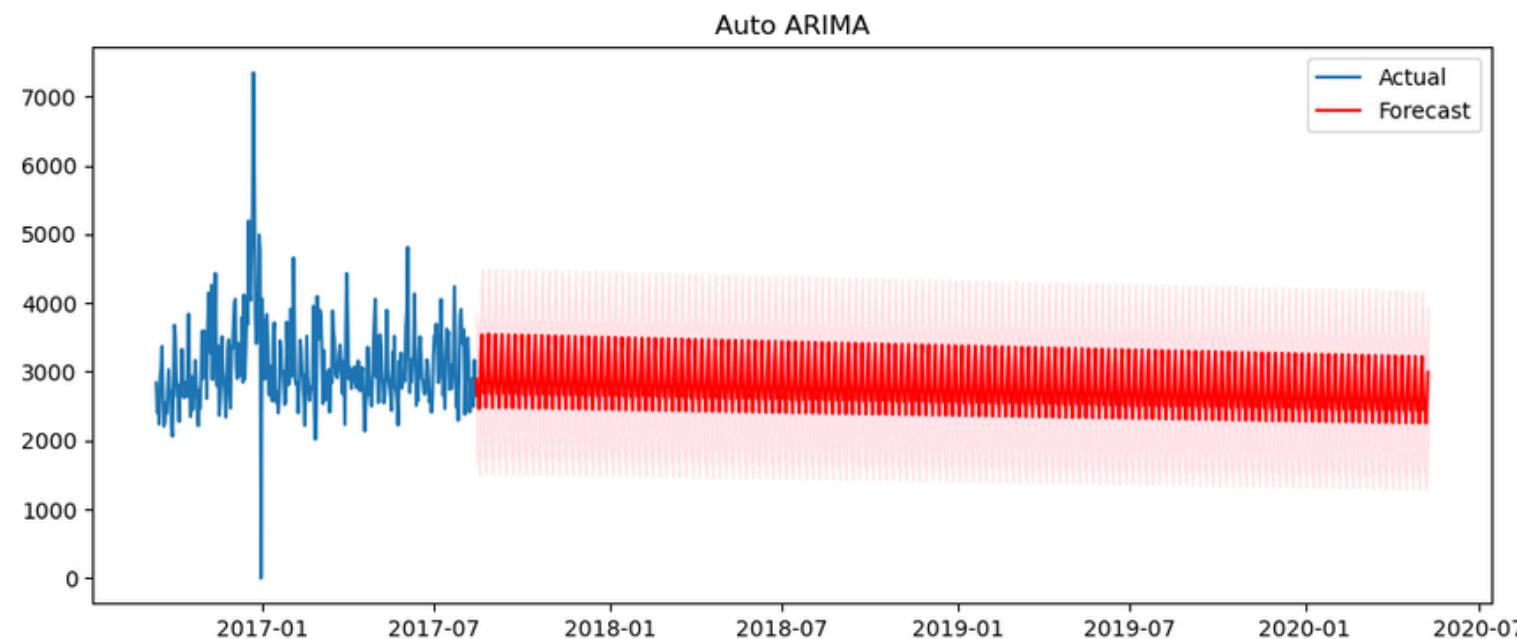
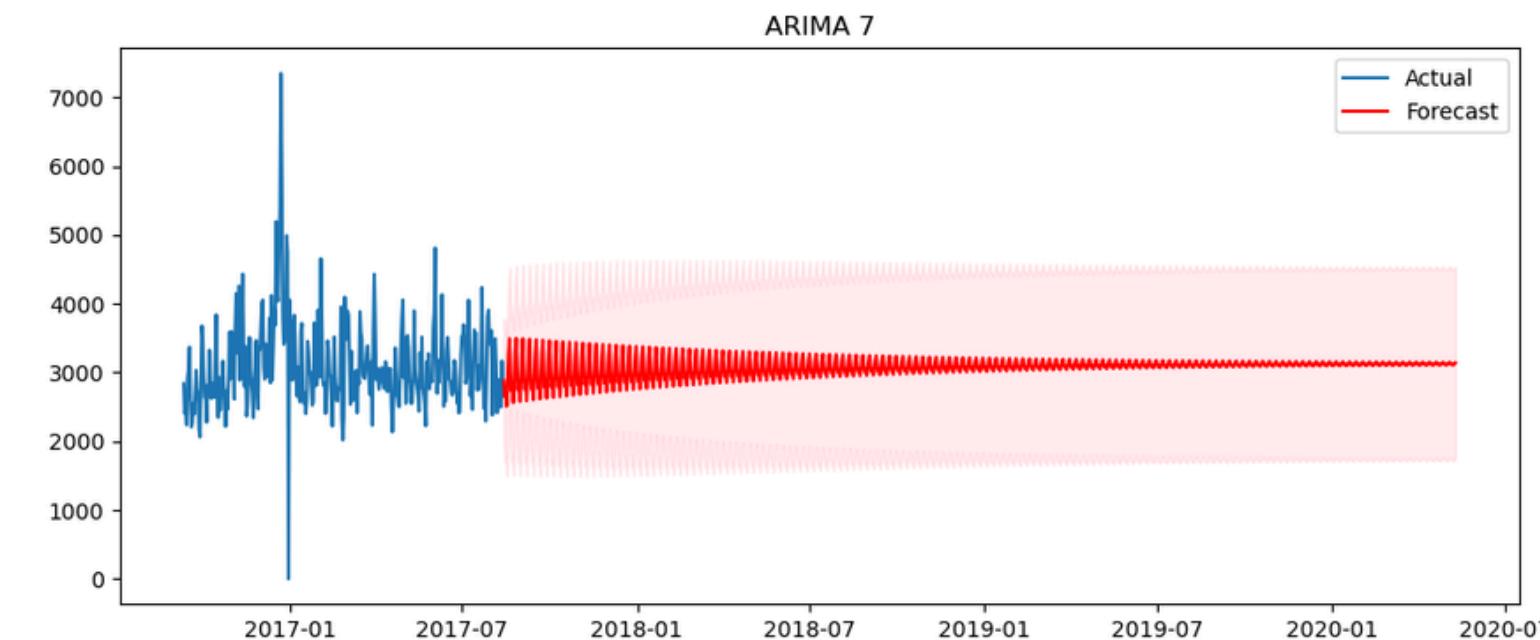
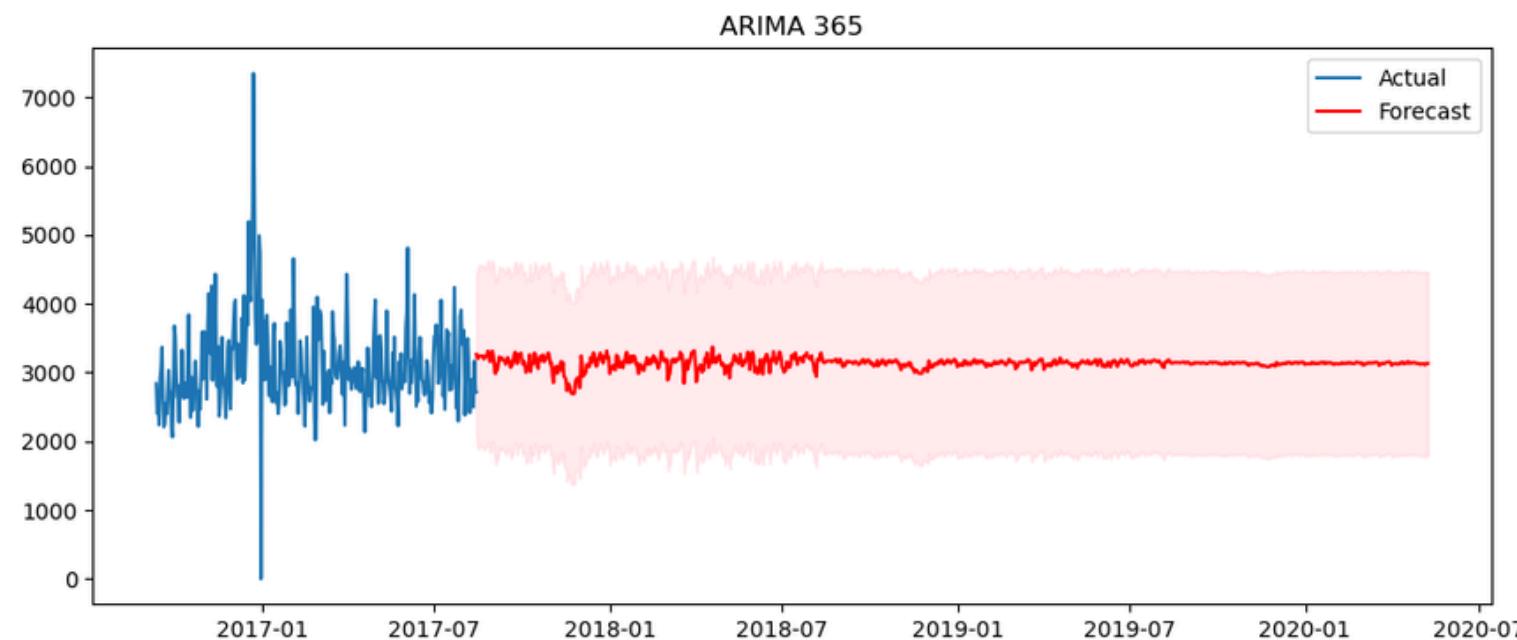
ARIMA & LSTM Model Comparison

Comparison MAE , MAPE & RSME

| Category | MAPE | MAE | RMSE |
|--------------------|-----------|--------------|--------------|
| ARIMA Daily 365 | 15.990586 | 509.514655 | 701.396566 |
| Arima Daily 7 | 13.018163 | 434.049301 | 635.692635 |
| Auto Arima Daily | 18.094961 | 589.298621 | 774.275163 |
| LSTM Daily | 7.800415 | 226.391062 | 285.350773 |
| ARIMA Monthly 12 | 19.119256 | 12554.115774 | 18900.419547 |
| Arima Monthly 7 | 17.378982 | 11444.167712 | 18345.308551 |
| Auto Arima Monthly | 14.976582 | 9694.803512 | 16600.808415 |
| LSTM Monthly | 24.499753 | 15621.962891 | 26776.410614 |

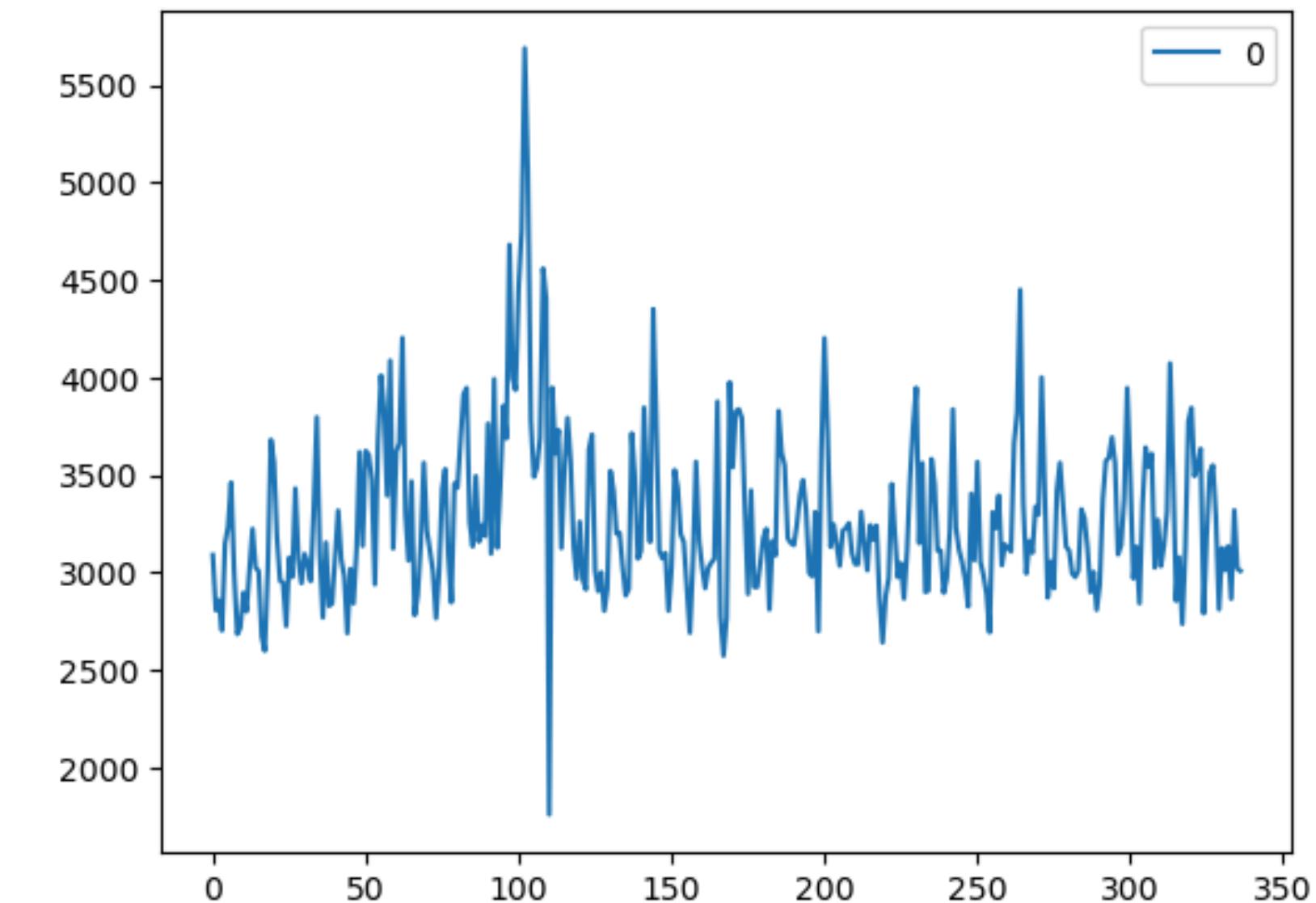
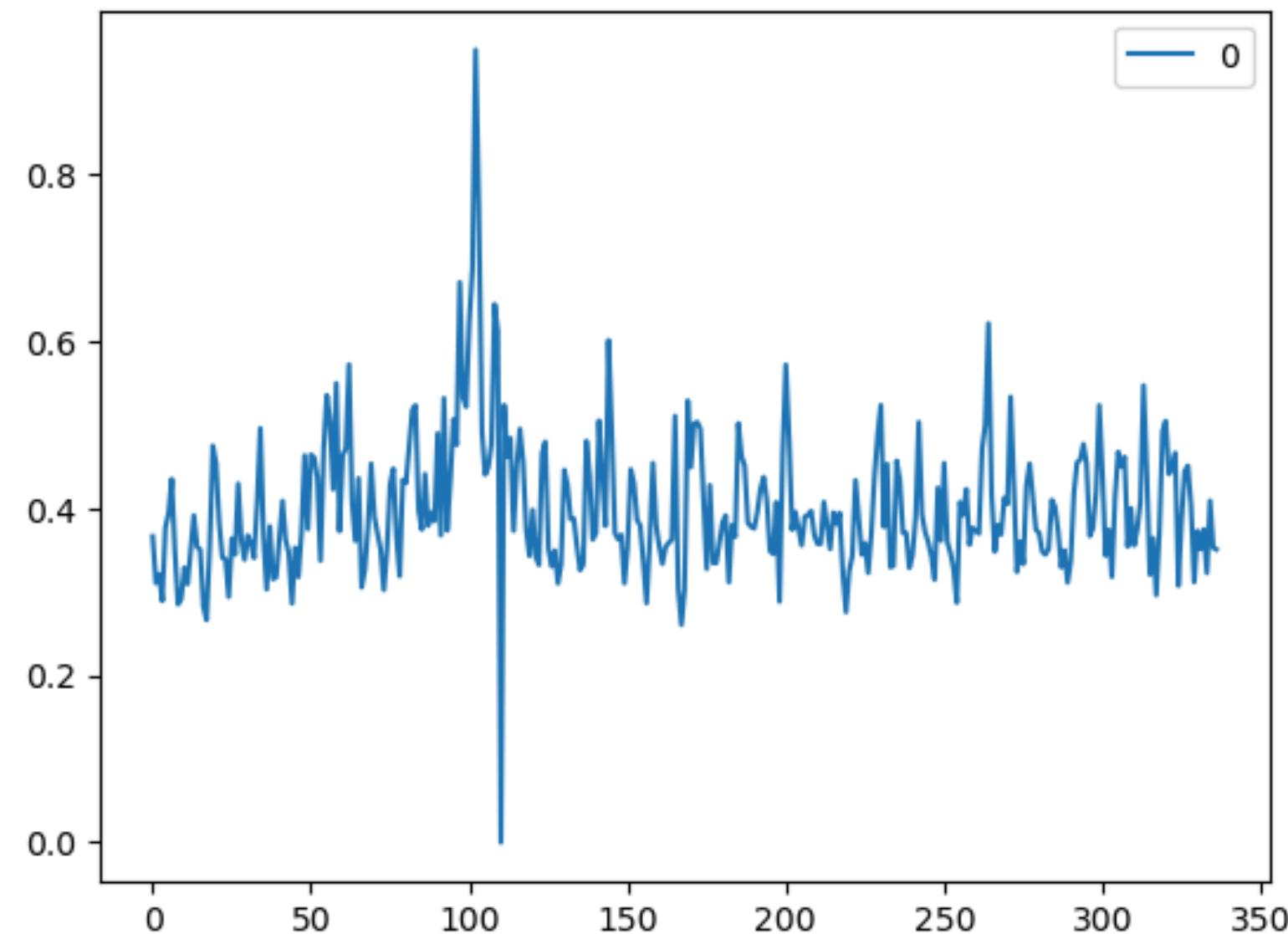
Model Development

ARIMA Model (Forecasting Opsional)



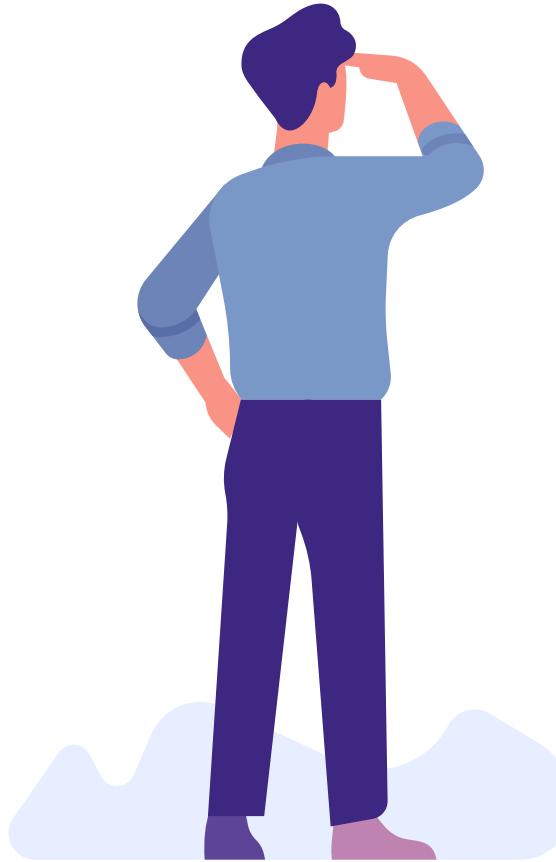
Model Development

LSTM (Forecasting The Best Model)



Future Improvement

1. Model dapat dikembangkan untuk menganalisa kategori lainnya
2. Hasil model dapat dikembangkan untuk insight bisnis seperti analisa stock mingguan, prediksi lonjakan penjualan, prediksi penurunan
3. Model dapat diintegrasikan ke sistem untuk notifikasi.



Conclusion

Model ARIMA lebih cocok untuk data time series jangka pendek
sedangkan untuk Model LSTM lebih cocok untuk data time series yang kompleks.

