

PREDIKSI NILAI INDEX LQ45

(oleh: Eric Arinoto)

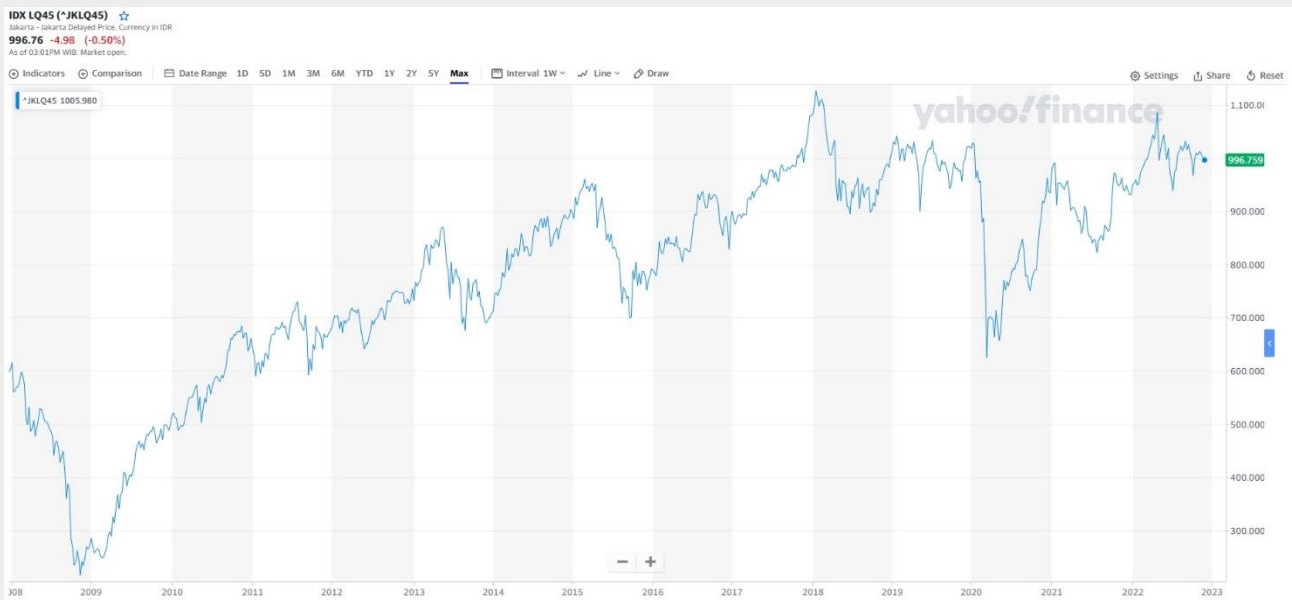
https://github.com/ercainz/lq45_prediction

1. PERMASALAHAN BISNIS

Pada dunia pasar modal, tentunya banyak orang yang ingin mendapatkan keuntungan yang maksimal. Untuk mencapai tujuan tersebut, sebenarnya ada banyak metode analisa yang dapat dilakukan. Namun untuk mempelajari dan menerapkannya diperlukan waktu dan ketekunan yang konsisten setiap harinya. Hal itu relatif sulit bagi sebagian orang, terlebih lagi bagi mereka yang mempunyai rutinitas yang padat.

Pada project ini, dimaksudkan untuk mengatasi problematik tersebut. Project ini akan mencoba melakukan prediksi harga penutupan salah satu indeks saham yang cukup populer di kalangan dunia pasar modal, yaitu Indeks LQ45.

Nilai indeks LQ45 yang akan diprediksi adalah harga penutupan pada 3 hari ke depan. Pemilihan angka 3 hari didapatkan dari hasil analisa dari beberapa analis saham, yang mengatakan bahwa kejadian di T0 idealnya akan terasa efeknya hingga 3-4 hari ke depan.



Gambar 1 : pergerakan harga penutupan Index LQ45 (2008-2022)

2. OBYEKTIF DAN METRIKS BISNIS

Project ini dimaksudkan agar dengan mengetahui harga penutupan index LQ45 pada 3 hari ke depan, User dapat melakukan tindakan yang tepat (beli atau jual) untuk memaksimalkan keuntungan mereka.

Tingkat keakuratan dari angka hasil prediksi tidak terlalu diutamakan dalam project ini. Namun lebih diutamakan untuk dapat mengetahui arah pergerakan nilai indeks tersebut (positif atau negatif) dengan tepat.

3. SOLUSI MACHINE LEARNING

3.1 PEMILIHAN MODEL

Dari Gambar 1, dapat dilihat bahwa data yang akan diprediksi berupa deret waktu. Salah satu model machine learning yang umum digunakan untuk melakukan pemodelan regresi deret waktu adalah ARIMA. ARIMA merupakan jenis model regresi univariate, sehingga yang tidak memerlukan fitur tambahan apapun untuk menghasilkan prediksi. ARIMA memiliki 3 parameter dasar, yaitu AR -

Autoregressive (**p**), I - Integrated (**d**), MA - Moving Average (**q**). Model ARIMA memiliki beberapa pengembangan seiring kemajuan jaman, yaitu SARIMA, ARIMAX, dan SARIMAX.

(S)ARIMA merupakan pengembangan model ARIMA dengan tambahan parameter seasonal yang dapat mendeteksi adanya efek musiman untuk periode waktu yang lebih besar. Misalnya, untuk data dengan rentang waktu bulanan, efek musiman dapat melihat pola seasonal tahunan (per 12 bulan). SARIMA memiliki tambahan 3 parameter tambahan selain seasonal (**S**), yaitu Seasonal-Autoregressive (**P**), Seasonal-Integrated (**D**), Seasonal-Moving Average (**Q**).

ARIMA(X) merupakan pengembangan model ARIMA yang menggunakan tambahan fitur-fitur eksogen untuk menghasilkan prediksi. Hal ini dipercaya dapat meningkatkan performa dari model ARIMA biasa.

(S)ARIMA(X) merupakan gabungan dari model SARIMA dan ARIMAX. Model inilah yang akan digunakan pada project ini.

3.2 PARAMETER MODEL

Model SARIMAX yang merupakan gabungan dari model SARIMA dan ARIMAX memiliki 7 parameter dan merupakan jenis model regresi multivariate yang artinya bisa memiliki fitur-fitur eksogen (X) yang lebih dari 1.

Untuk parameter p , d , q , P , D , Q akan dipilih dengan menggunakan pustaka [pmdarima](#) yang dapat secara otomatis mencari nilai terbaik dari keenam parameter tersebut. Sedangkan parameter S , akan dipilih angka 20. Pemilihan ini dimaksudkan untuk mendapatkan trend musiman secara bulanan (1 bulan rata-rata memiliki 20 hari kerja).

Fitur-fitur (X) yang akan digunakan untuk pelatihan model yaitu:

1. **JCI** - Harga penutupan Indeks IHSG pada T_0 .
2. **IDX30** - Harga penutupan Indeks IDX-30 pada T_0 .
3. **EIDO** - Harga penutupan Indeks EIDO pada T_0 .
4. **SPY** - Harga penutupan Indeks S&P 500 pada T_0 .
5. **DOM_B** - Jumlah nominal transaksi pembelian domestik pada T_0 .
6. **DOM_S** - Jumlah nominal transaksi penjualan domestik pada T_0 .
7. **FOR_B** - Jumlah nominal transaksi pembelian asing pada T_0 .
8. **FOR_S** - Jumlah nominal transaksi penjualan asing pada T_0 .

Selain dari 8 fitur yang akan didapat dari data mentah, juga akan ditambahkan 4 fitur tambahan lagi yang merupakan hasil penggabungan dari beberapa fitur yang sudah ada, yaitu:

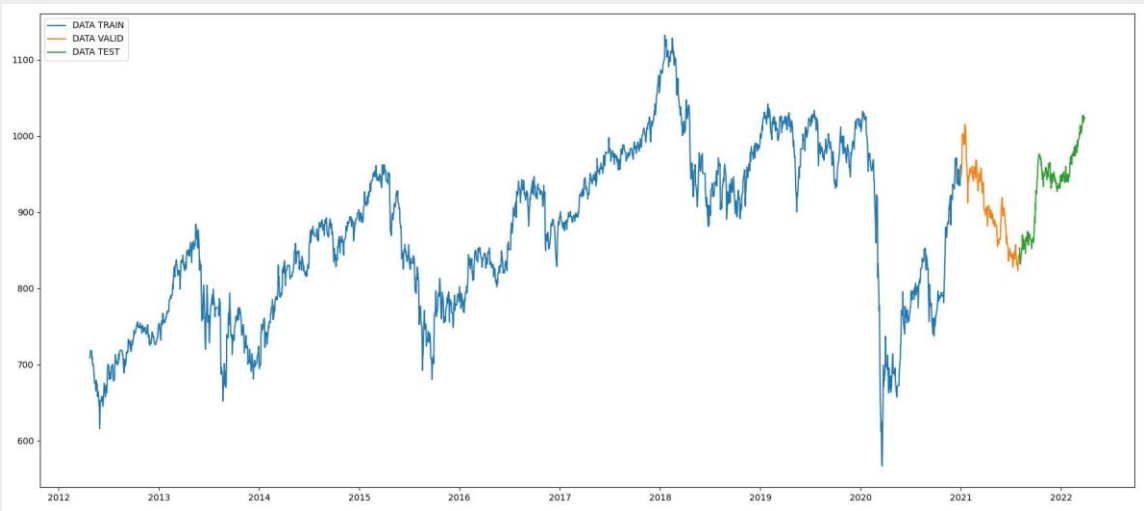
1. **DOM_TOTAL** - merupakan kolom **DOM_B** dijumlahkan dengan kolom **DOM_S**.
2. **DOM_NET** - merupakan kolom **DOM_B** dikurangkan dengan kolom **DOM_S**.
3. **FOR_TOTAL** - merupakan kolom **FOR_B** dijumlahkan dengan kolom **FOR_S**.
4. **FOR_NET** - merupakan kolom **FOR_B** dikurangkan dengan kolom **FOR_S**.

Sedangkan **TARGET** (y) yang dicapai adalah harga penutupan index LQ45 pada $T+3$.

4. EKSPLORASI

Data yang akan digunakan untuk pelatihan model merupakan data deret waktu dengan interval harian (hanya senin hingga jumat), sejak April 2012 hingga Maret 2022. Yang akan dibagi dengan komposisi sebagai berikut:

- data train adalah data sejak awal hingga 31-12-2020.
- data validasi adalah data sejak 01-01-2021 hingga 31-07-2021.
- data test adalah data sejak 01-08-2021 hingga data terakhir.

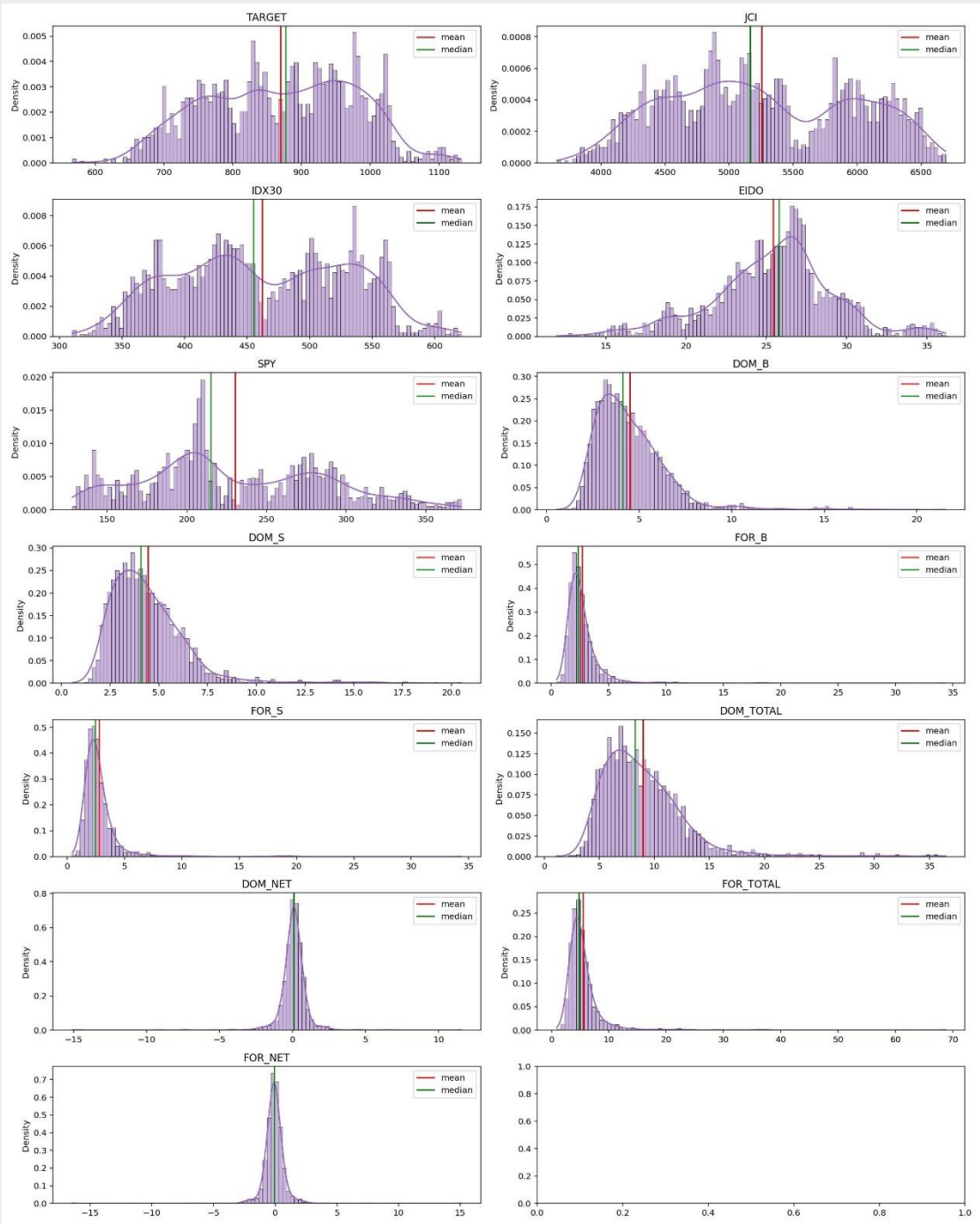


Gambar 2 : data target (y)

4.1 DISTRIBUSI DATA

Pada *Gambar 3* yang merupakan distribusi dari data train, dapat dilihat bahwa persebaran dari sebagian besar fitur tidak normal. Titik mean (garis merah) dan median (garis hijau) cukup jauh pada fitur JCI, IDX30, EIDO, SPY, DOM_B, DOM_S, FOR_B, FOR_S, DOM_TOTAL, dan FOR_TOTAL, bahkan pada TARGET.

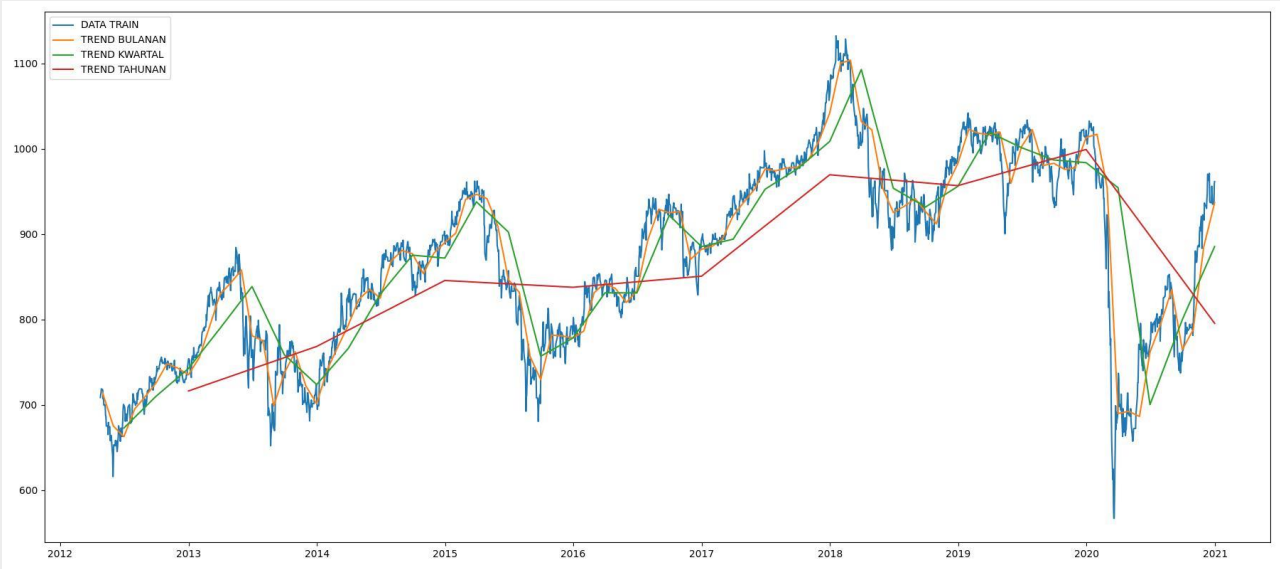
Dengan demikian perlu dilakukan beberapa langkah preprocessing data sebelum data dapat digunakan untuk pelatihan model.



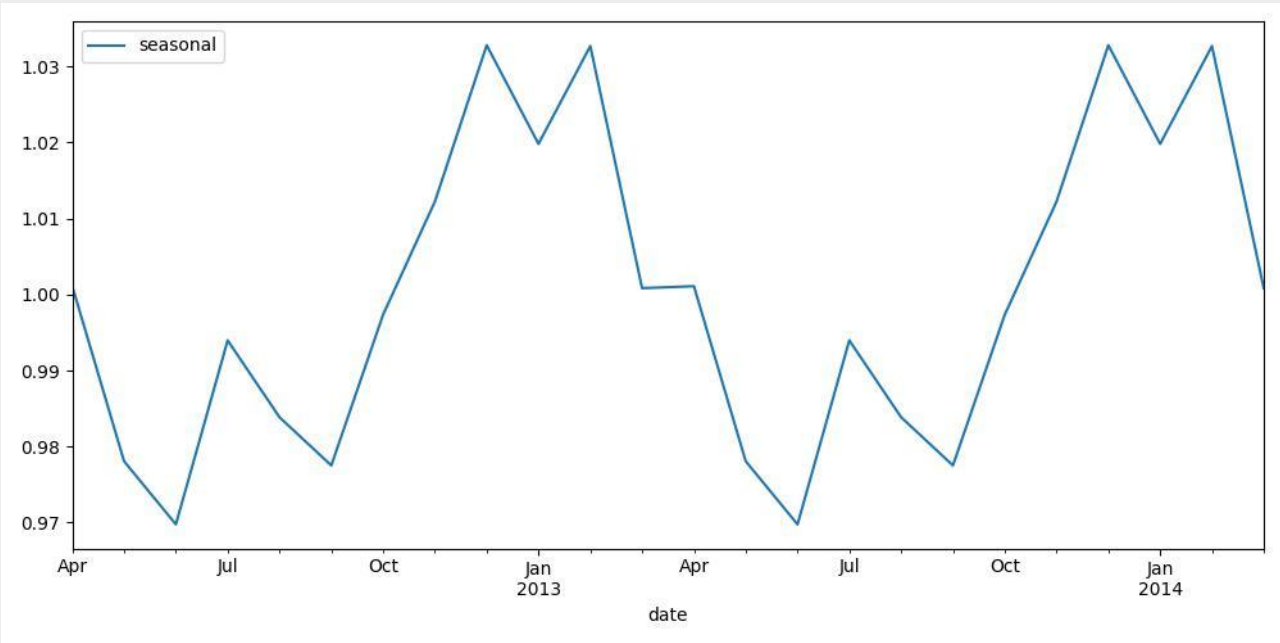
Gambar 3 : distribusi data

4.2 TREND DARI TARGET

Pada *Gambar 4* menunjukkan trend dari TARGET untuk interval bulanan (orange), interval kwartal (hijau) dan interval tahunan (merah). Untuk pelatihan model akan digunakan interval bulanan.



Gambar 4 : trend dari data target



Gambar 5 : trend musiman bulanan

Sedangkan pada *Gambar 5* merupakan trend musiman bulanan selama 2 tahun. Data musiman ini didapatkan dari proses dekomposisi seasonal. Data ini juga akan digunakan sebagai fitur tambahan untuk proses pelatihan model.

Lebih detail mengenai proses dekomposisi seasonal, dapat dibaca pada artikel [disini](#).

4.3 STATIONARY DATA

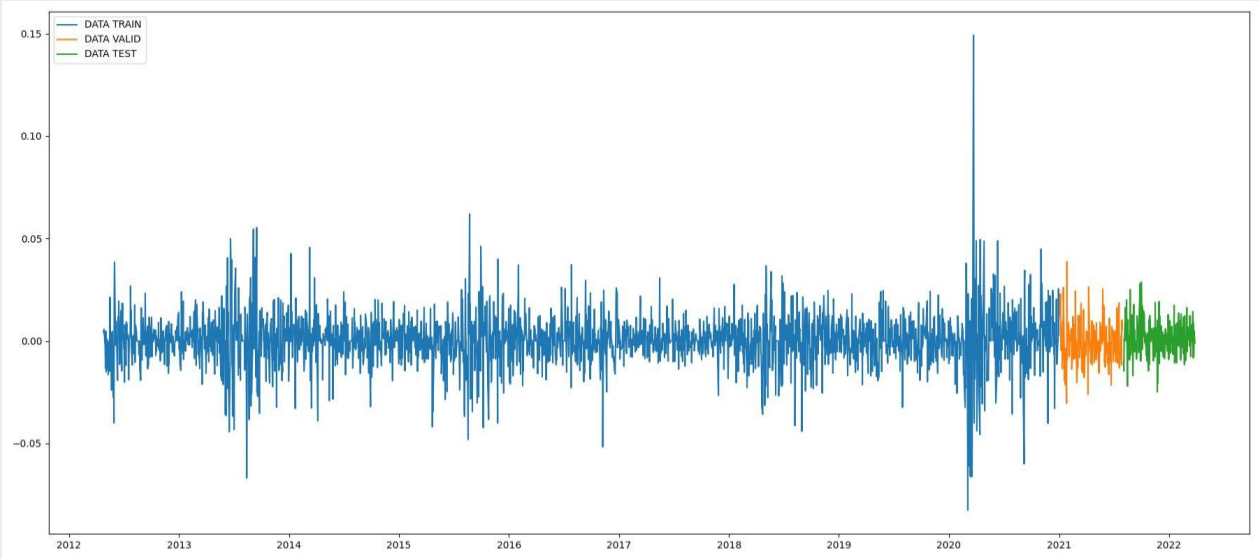
Model ARIMA dan turunannya mewajibkan data yang stasioner untuk digunakan saat pelatihan model. Dari smua fitur akan dilakukan pengujian stasioneritas menggunakan *Uji Augmented Dickey-Fuller (ADF Test)*. Data deret waktu bisa disebut stasioner apabila nilai p-value dibawah angka kritikal 0.05 dan sangat mendekati nol. Lebih detail mengenai ADF Test, dapat dibaca pada artikel [disini](#).

COLUMN	: target
1. ADF	: -2.443714416414124
2. P-Value	: 0.12977641072082524
3. Num Of Lags	: 14
4. Num Of Observations Used	: 2253
5. Critical Values 1%	: -3.433255799449242
6. Critical Values 5%	: -2.862823704909085
7. Critical Values 10%	: -2.56745337629031
RESULT	: NON-STATIONARY

Gambar 6 : uji stasioner dengan ADF Test

Gambar 6 adalah hasil test stasioner dari TARGET pada data train. Dari hasil tersebut bisa disimpulkan bahwa data TARGET tidak stasioner (p-value masih lebih besar dari angka kritikal 0.05).

Untuk mengatasi hal ini, perlu dilakukan proses differencing. Gambar 7 adalah data TARGET setelah dilakukan proses differencing lag 1. Sedangkan Gambar 8 adalah hasil test setelah proses differencing lag 1.



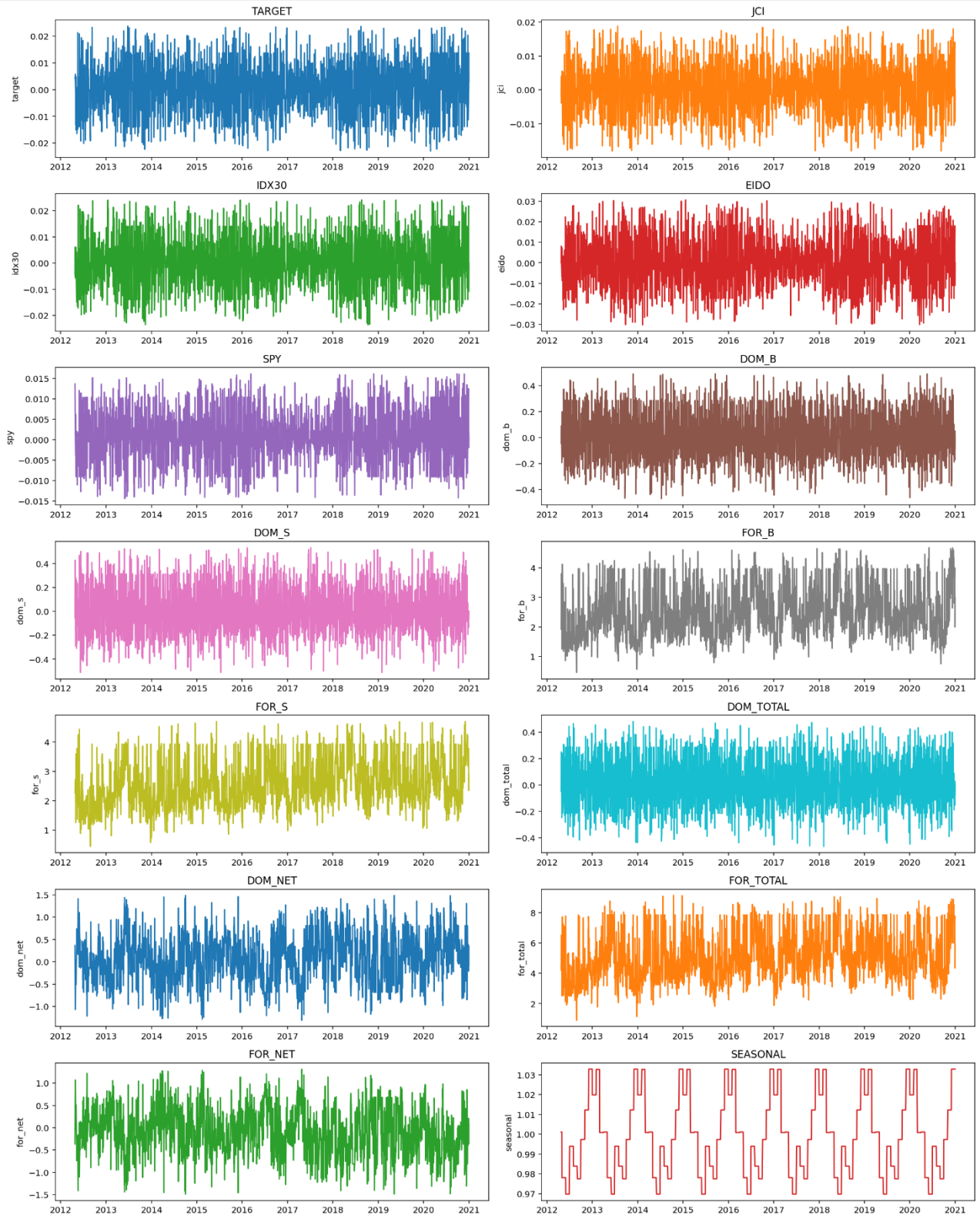
Gambar 7 : data target (setelah proses differencing)

COLUMN	: target
1. ADF	: -14.584418343546046
2. P-Value	: 4.415433212973363e-27
3. Num Of Lags	: 13
4. Num Of Observations Used	: 2253
5. Critical Values 1%	: -3.433255799449242
6. Critical Values 5%	: -2.862823704909085
7. Critical Values 10%	: -2.56745337629031
RESULT	: STATIONARY

Gambar 8 : uji stasioner dengan ADF Test (setelah proses differencing)

Dari hasil tersebut bisa disimpulkan bahwa data TARGET saat ini sudah stasioner (p-value sangat mendekati angka nol).

Pengujian stasioneritas dengan cara yang sama juga dilakukan pada masing-masing fitur. Gambar 9 adalah data TARGET dan masing-masing fitur setelah dilakukan proses differencing lag 1.

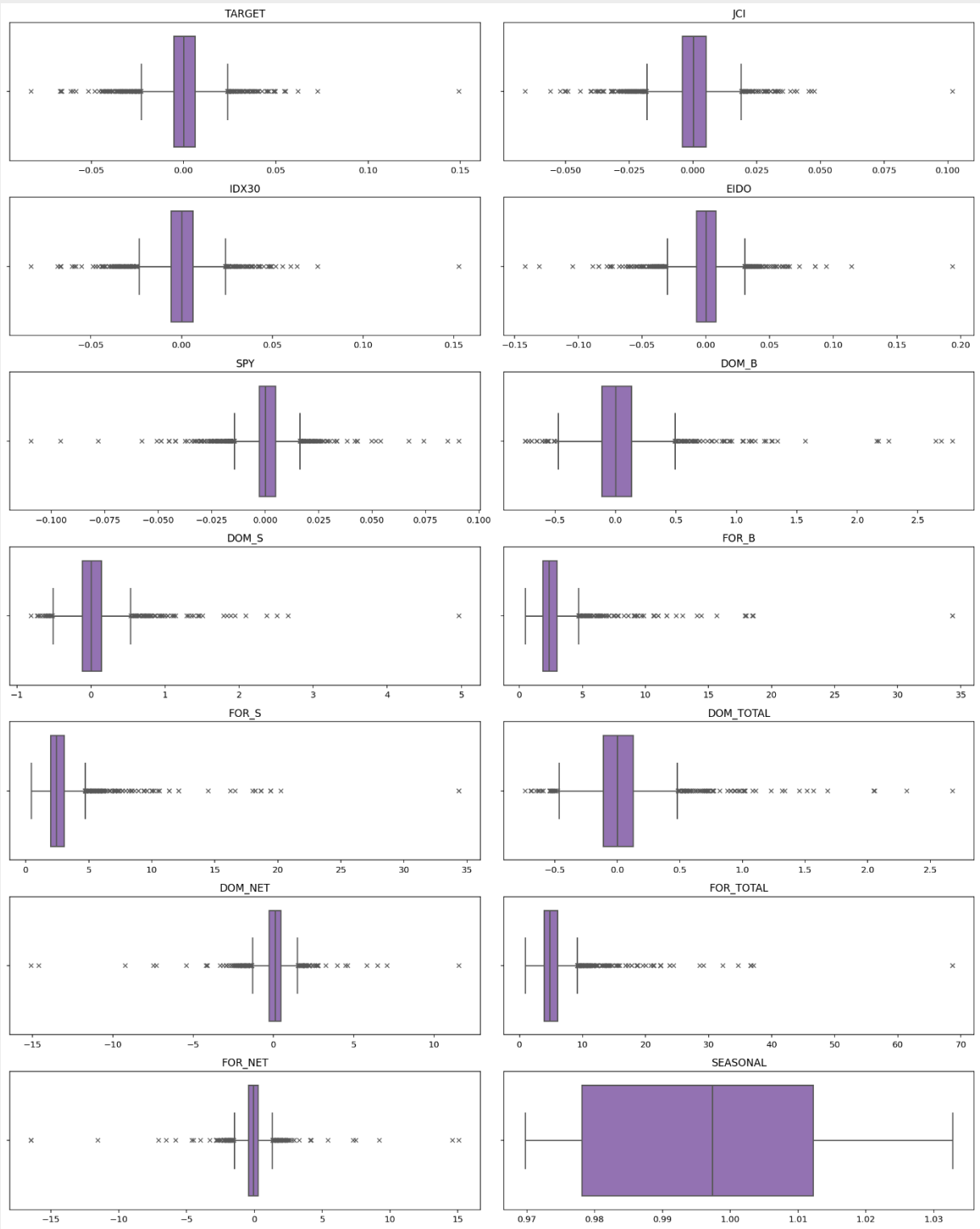


Gambar 9 : TARGET dan masing-masing fitur (setelah proses differencing)

4.4 PENANGANAN OUTLIER

Pendeteksian data outlier pada semua kolom menggunakan metode 1.5 x IQR, serta melakukan imputasi dengan nilai percentile 10% untuk outlier bawah, dan percentile 90% untuk outlier atas. Imputasi tidak menggunakan nilai mean maupun median. Hal ini dikarenakan data dengan jarak yang jauh dari nilai mean adalah data yang justru ingin dipelajari oleh model. Sehingga sebisa mungkin data-data tersebut tetap dapat menunjukkan jarak yang signifikan terhadap nilai mean.

Gambar 10 merupakan boxplot sebelum dilakukan imputasi.



Gambar 10 : boxplot TARGET dan masing-masing fitur

4.5 EKSPLORASI PEMODELAN

Proses standardisasi dilakukan sebelum proses pelatihan model. Proses standardisasi ini dilakukan pada semua fitur saja, sedangkan data target akan tetap dibiarkan apa adanya.

Proses pelatihan model menggunakan pustaka *pmdarima* sehingga dapat secara otomatis mencari nilai terbaik dari semua parameter model, kecuali parameter seasonal (**S**) menggunakan angka 20.

Berikut pada *Gambar 11.a* dan *Gambar 11.b* adalah hasil dari pelatihan model SARIMAX. Hasil pelatihan menunjukkan bahwa parameter terbaik untuk permasalahan ini adalah ARIMA (2,0,2) (0,0,1) [20] yang artinya:

- **Auto Regressive:** p=2
- **Integrated:** d=0
- **Moving Average:** q=2
- **Seasonal Auto Regressive:** P=0
- **Seasonal Integrated:** D=0
- **Seasonal Moving Average:** Q=1
- **Seasonal:** S=20

```
Performing stepwise search to minimize aic
ARIMA(2,0,2)(1,0,1)[20] intercept : AIC=-14750.247, Time=17.05 sec
ARIMA(0,0,0)(0,0,0)[20] intercept : AIC=-14731.627, Time=1.28 sec
ARIMA(1,0,0)(1,0,0)[20] intercept : AIC=-14731.889, Time=12.15 sec
ARIMA(0,0,1)(0,0,1)[20] intercept : AIC=-14731.980, Time=6.69 sec
ARIMA(0,0,0)(0,0,0)[20] : AIC=-14730.036, Time=0.99 sec
ARIMA(2,0,2)(0,0,1)[20] intercept : AIC=-14752.547, Time=28.10 sec
ARIMA(2,0,2)(0,0,0)[20] intercept : AIC=-14750.341, Time=5.19 sec
ARIMA(2,0,2)(0,0,2)[20] intercept : AIC=-14750.267, Time=45.88 sec
ARIMA(2,0,2)(1,0,0)[20] intercept : AIC=-14752.394, Time=22.55 sec
ARIMA(2,0,2)(1,0,2)[20] intercept : AIC=-14748.328, Time=47.06 sec
ARIMA(1,0,2)(0,0,1)[20] intercept : AIC=-14747.264, Time=14.70 sec
ARIMA(2,0,1)(0,0,1)[20] intercept : AIC=-14744.869, Time=20.39 sec
ARIMA(3,0,2)(0,0,1)[20] intercept : AIC=-14750.479, Time=13.81 sec
ARIMA(2,0,3)(0,0,1)[20] intercept : AIC=-14749.851, Time=15.20 sec
ARIMA(1,0,1)(0,0,1)[20] intercept : AIC=-14730.042, Time=16.06 sec
ARIMA(1,0,3)(0,0,1)[20] intercept : AIC=-14745.101, Time=9.47 sec
ARIMA(3,0,1)(0,0,1)[20] intercept : AIC=-14742.820, Time=16.84 sec
ARIMA(3,0,3)(0,0,1)[20] intercept : AIC=-14743.453, Time=29.02 sec
ARIMA(2,0,2)(0,0,1)[20] : AIC=-14749.166, Time=32.54 sec

Best model: ARIMA(2,0,2)(0,0,1)[20] intercept
Total fit time: 355.015 seconds
```

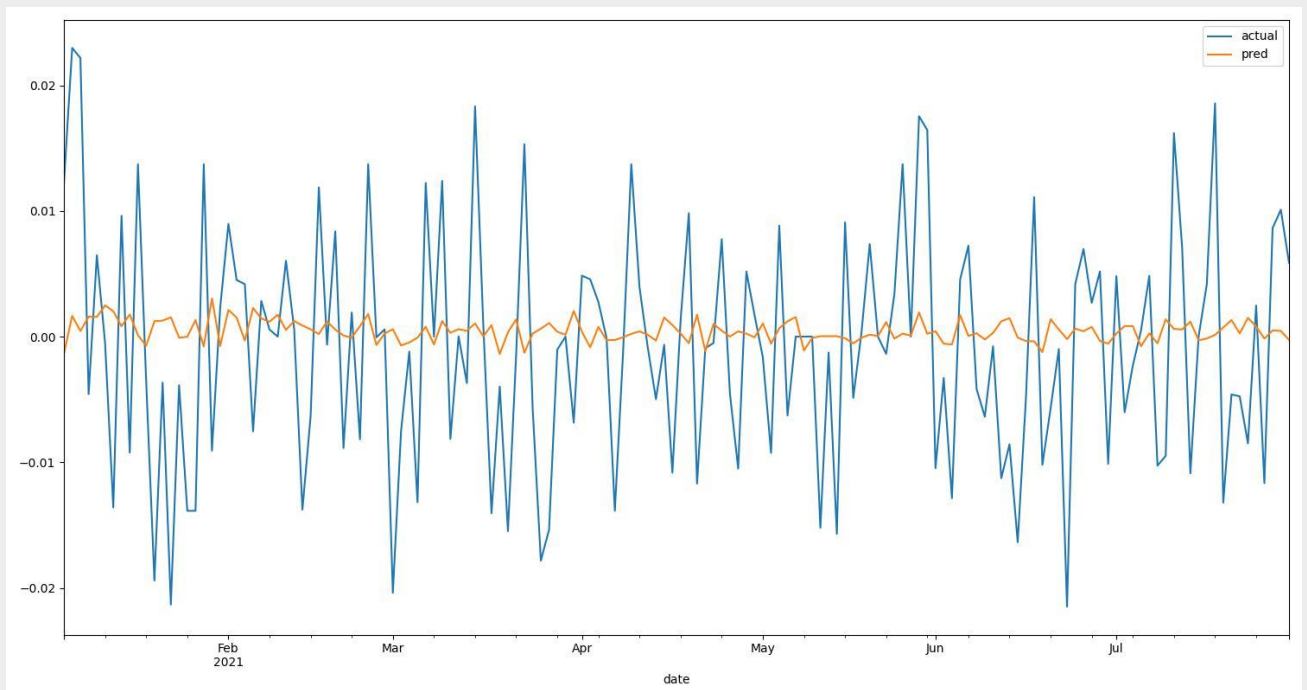
Gambar 11.a : hasil pelatihan model SARIMAX (a)

```
SARIMAX Results
=====
Dep. Variable: y No. Observations: 2267
Model: SARIMAX(2, 0, 2)x(0, 0, [1], 20) Log Likelihood 7396.274
Date: Tue, 22 Nov 2022 AIC -14752.547
Time: 14:54:48 BIC -14638.023
Sample: 04-25-2012 HQIC -14710.762
- 12-31-2020
Covariance Type: opg
=====
coef std err z P>|z| [0.025 0.975]
-----
intercept 0.0001 8.17e-05 1.765 0.078 -1.6e-05 0.000
jci 0.0001 0.001 0.199 0.843 -0.001 0.002
idx30 -0.0006 0.001 -0.870 0.384 -0.002 0.001
eido 0.0003 0.000 0.973 0.330 -0.000 0.001
spy 0.0004 0.000 1.953 0.051 -1.47e-06 0.001
dom_b 0.0006 0.001 0.893 0.372 -0.001 0.002
dom_s -0.0009 0.001 -1.343 0.179 -0.002 0.000
for_b -0.0015 0.001 -1.271 0.204 -0.004 0.001
for_s -0.0018 0.001 -1.564 0.118 -0.004 0.000
dom_total 0.0002 0.001 0.230 0.818 -0.002 0.002
dom_net -0.0041 0.002 -2.508 0.012 -0.007 -0.001
for_total 0.0029 0.002 1.518 0.129 -0.001 0.007
for_net -0.0038 0.002 -2.225 0.026 -0.007 -0.000
seasonal 0.0003 0.000 1.628 0.103 -5.4e-05 0.001
ar.L1 0.0451 0.132 0.342 0.732 -0.213 0.303
ar.L2 0.5529 0.128 4.315 0.000 0.302 0.804
ma.L1 -0.0493 0.125 -0.396 0.692 -0.293 0.195
ma.L2 -0.6422 0.119 -5.406 0.000 -0.875 -0.409
ma.S.L20 0.0416 0.020 2.047 0.041 0.002 0.081
sigma2 8.574e-05 2.73e-06 31.365 0.000 8.04e-05 9.11e-05
=====
Ljung-Box (L1) (Q): 0.00 Jarque-Bera (JB): 4.92
Prob(Q): 0.98 Prob(JB): 0.09
Heteroskedasticity (H): 1.22 Skew: -0.02
Prob(H) (two-sided): 0.01 Kurtosis: 2.77
=====

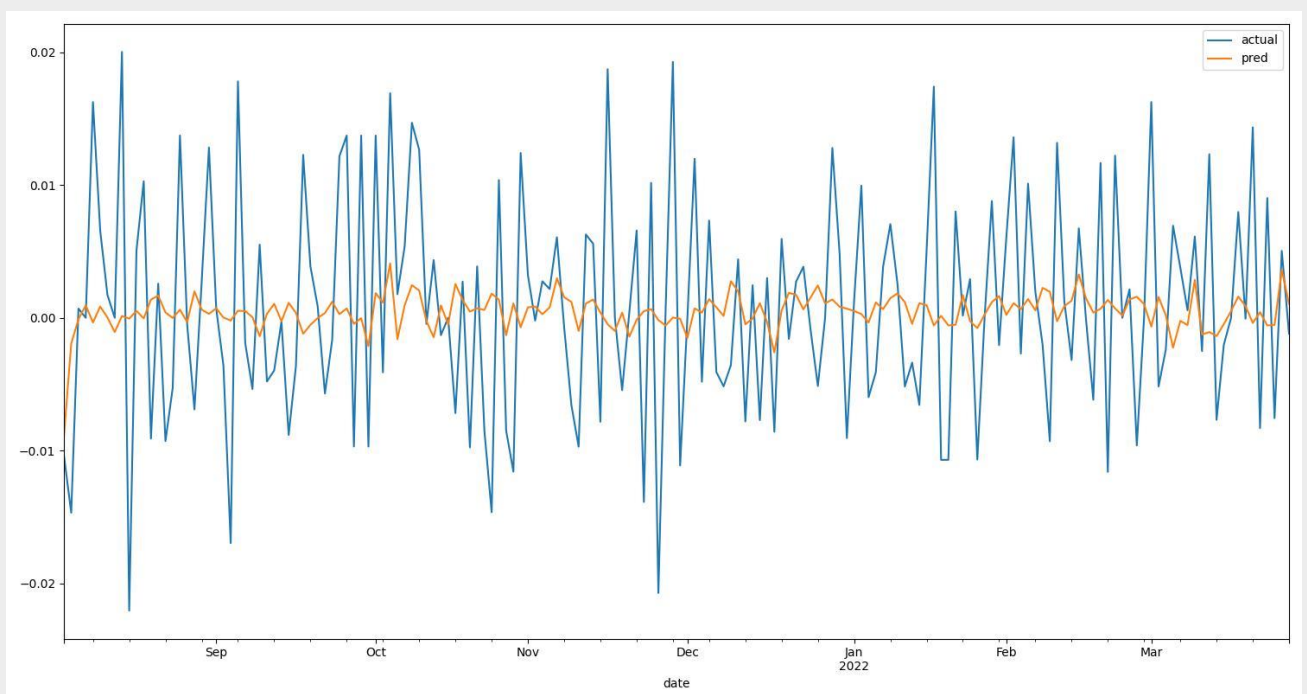
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Gambar 11.b : hasil pelatihan model SARIMAX (b)

Dengan model tersebut dilakukan prediksi terhadap data validasi dan data test. Berikut pada *Gambar 12* dan *Gambar 13* adalah hasilnya.



Gambar 12 : hasil prediksi pada data validasi



Gambar 13 : hasil prediksi pada data test

Dari *Gambar 12* dan *Gambar 13* dapat dilihat bahwa performa dari model yang sudah dilatih cukup buruk. Hasil prediksi terlalu takut untuk bergerak terlalu jauh dari titik nol.

Selanjutnya akan dicoba kembali melakukan pelatihan model menggunakan data yang tidak distasionerkan. Diharapkan akan mendapatkan performa model yang lebih baik. Berikut adalah hasil parameter model terbaik yang didapatkan adalah ARIMA (1,0,3) (2,0,0) [20] yang artinya:

- **Auto Regressive:** $p=1$
- **Integrated:** $d=0$
- **Moving Average:** $q=3$
- **Seasonal Auto Regressive:** $P=2$
- **Seasonal Integrated:** $D=0$
- **Seasonal Moving Average:** $Q=0$
- **Seasonal:** $S=20$

Gambar 14.a dan *Gambar 14.b* adalah hasil dari pelatihan model SARIMAX menggunakan data yang tidak distasionerkan.

```

Performing stepwise search to minimize aic
ARIMA(2,0,2)(1,0,1)[20] intercept : AIC=17688.002, Time=24.19 sec
ARIMA(0,0,0)(0,0,0)[20] intercept : AIC=20289.611, Time=0.20 sec
ARIMA(1,0,0)(1,0,0)[20] intercept : AIC=inf, Time=31.90 sec
ARIMA(0,0,1)(0,0,1)[20] intercept : AIC=18953.669, Time=16.53 sec
ARIMA(0,0,0)(0,0,0)[20] intercept : AIC=37166.270, Time=1.92 sec
ARIMA(2,0,2)(0,0,1)[20] intercept : AIC=17732.221, Time=21.57 sec
ARIMA(2,0,2)(1,0,0)[20] intercept : AIC=inf, Time=14.15 sec
ARIMA(2,0,2)(2,0,1)[20] intercept : AIC=17639.473, Time=75.40 sec
ARIMA(2,0,2)(2,0,0)[20] intercept : AIC=17720.667, Time=80.30 sec
ARIMA(2,0,2)(2,0,2)[20] intercept : AIC=17711.631, Time=81.74 sec
ARIMA(2,0,2)(1,0,2)[20] intercept : AIC=inf, Time=nan sec
ARIMA(1,0,2)(2,0,1)[20] intercept : AIC=17568.600, Time=73.02 sec
ARIMA(1,0,2)(1,0,1)[20] intercept : AIC=17647.508, Time=23.83 sec
ARIMA(1,0,2)(2,0,0)[20] intercept : AIC=17669.294, Time=77.89 sec
ARIMA(1,0,2)(2,0,2)[20] intercept : AIC=17632.740, Time=85.08 sec
ARIMA(1,0,2)(1,0,0)[20] intercept : AIC=inf, Time=nan sec
ARIMA(1,0,2)(1,0,2)[20] intercept : AIC=inf, Time=157.94 sec
ARIMA(0,0,2)(2,0,1)[20] intercept : AIC=17738.936, Time=62.22 sec
ARIMA(1,0,1)(2,0,1)[20] intercept : AIC=18201.825, Time=73.25 sec
ARIMA(1,0,3)(2,0,1)[20] intercept : AIC=17503.652, Time=68.56 sec
ARIMA(1,0,3)(1,0,1)[20] intercept : AIC=17520.201, Time=21.61 sec
ARIMA(1,0,3)(2,0,0)[20] intercept : AIC=17491.539, Time=69.21 sec
ARIMA(1,0,3)(1,0,0)[20] intercept : AIC=inf, Time=42.40 sec
ARIMA(0,0,3)(2,0,0)[20] intercept : AIC=17584.103, Time=58.84 sec
ARIMA(2,0,3)(2,0,0)[20] intercept : AIC=17504.479, Time=76.64 sec
ARIMA(1,0,4)(2,0,0)[20] intercept : AIC=17495.384, Time=71.99 sec
ARIMA(0,0,2)(2,0,0)[20] intercept : AIC=17788.981, Time=57.27 sec
ARIMA(0,0,4)(2,0,0)[20] intercept : AIC=17651.009, Time=63.65 sec
ARIMA(2,0,4)(2,0,0)[20] intercept : AIC=17510.844, Time=82.37 sec
ARIMA(1,0,3)(2,0,0)[20] intercept : AIC=18313.427, Time=67.56 sec

Best model: ARIMA(1,0,3)(2,0,0)[20] intercept
Total fit time: 1701.858 seconds

```

Gambar 14.a : hasil pelatihan model SARIMAX (a)

```

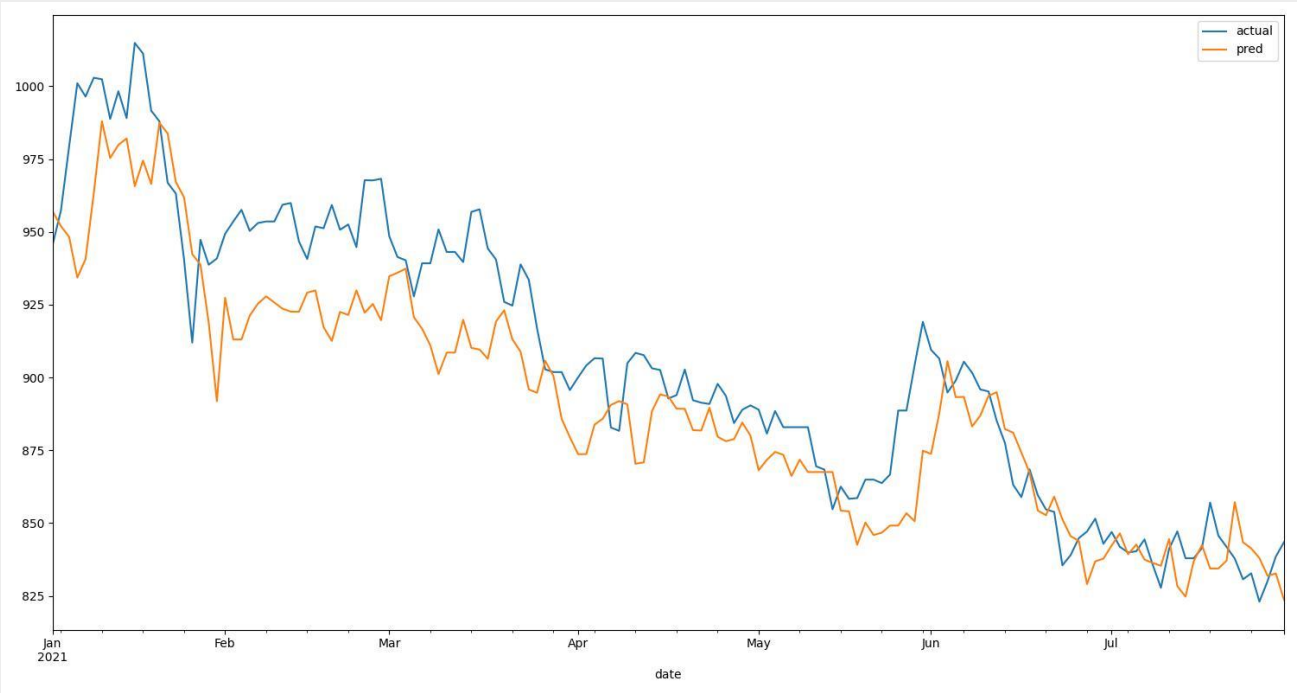
===== SARIMAX Results =====
Dep. Variable: y No. Observations: 2268
Model: SARIMAX(1, 0, 3)x(2, 0, [], 20) Log Likelihood: -8724.769
Date: Tue, 22 Nov 2022 AIC: 17491.539
Time: 14:51:00 BIC: 17611.799
Sample: 04-24-2012 HQIC: 17535.415
- 12-31-2020
Covariance Type: opg
=====
coef std err z P>|z| [0.025 0.975]
-----
intercept 139.6547 24.188 5.774 0.000 92.247 187.062
jci 11.1982 6.123 1.829 0.067 -0.803 23.200
idx30 100.7340 5.776 17.440 0.000 89.413 112.054
eido 4.5863 0.898 5.105 0.000 2.825 6.347
spy -9.8684 1.528 -6.458 0.000 -12.863 -6.873
dom_b 3.0923 0.846 3.654 0.000 1.434 4.751
dom_s 1.4603 0.906 1.612 0.107 -0.316 3.236
for_b -1.3842 0.694 -1.995 0.046 -2.744 -0.024
for_s -5.0591 0.723 -6.994 0.000 -6.477 -3.641
dom_total -5.5922 1.130 -4.950 0.000 -7.806 -3.378
dom_net 1.3210 1.079 1.224 0.221 -0.794 3.436
for_total 5.9803 1.095 5.463 0.000 3.835 8.126
for_net -0.3854 1.097 -0.351 0.725 -2.535 1.765
seasonal 2.5756 0.834 3.089 0.002 0.941 4.210
ar.L1 0.8363 0.029 28.877 0.000 0.780 0.893
ma.L1 0.2138 0.036 5.938 0.000 0.143 0.284
ma.L2 0.1541 0.037 4.138 0.000 0.081 0.227
ma.L3 -0.6088 0.034 -17.728 0.000 -0.676 -0.542
ar.S.L20 0.0217 0.020 1.086 0.278 -0.017 0.061
ar.S.L40 -0.0024 0.021 -0.115 0.908 -0.044 0.039
sigma2 132.5220 2.629 50.402 0.000 127.369 137.675
=====
Ljung-Box (L1) (Q): 0.08 Jarque-Bera (JB): 1434.34
Prob(Q): 0.78 Prob(JB): 0.00
Heteroskedasticity (H): 1.67 Skew: -0.24
Prob(H) (two-sided): 0.00 Kurtosis: 6.87
=====

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

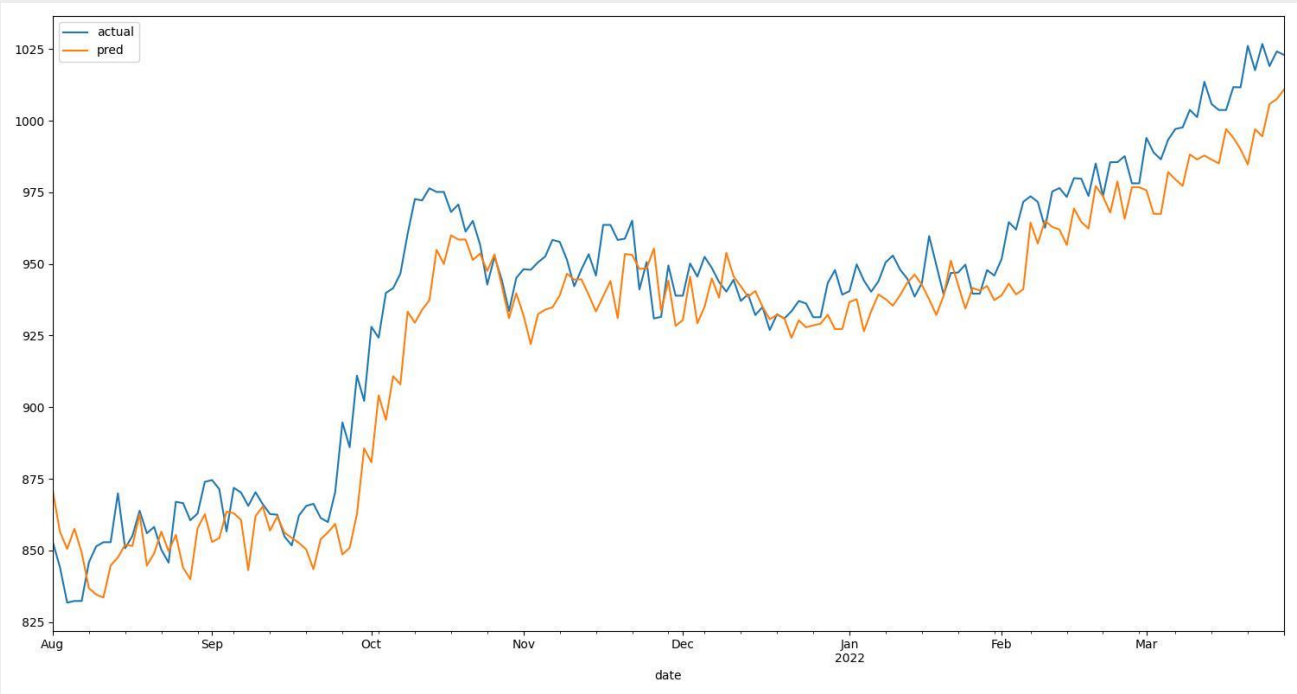
```

Gambar 14.a : hasil pelatihan model SARIMAX (b)

Dengan model tersebut dilakukan prediksi terhadap data validasi dan data test. Berikut pada *Gambar 15* dan *Gambar 16* adalah hasilnya.



Gambar 15 : hasil prediksi pada data validasi

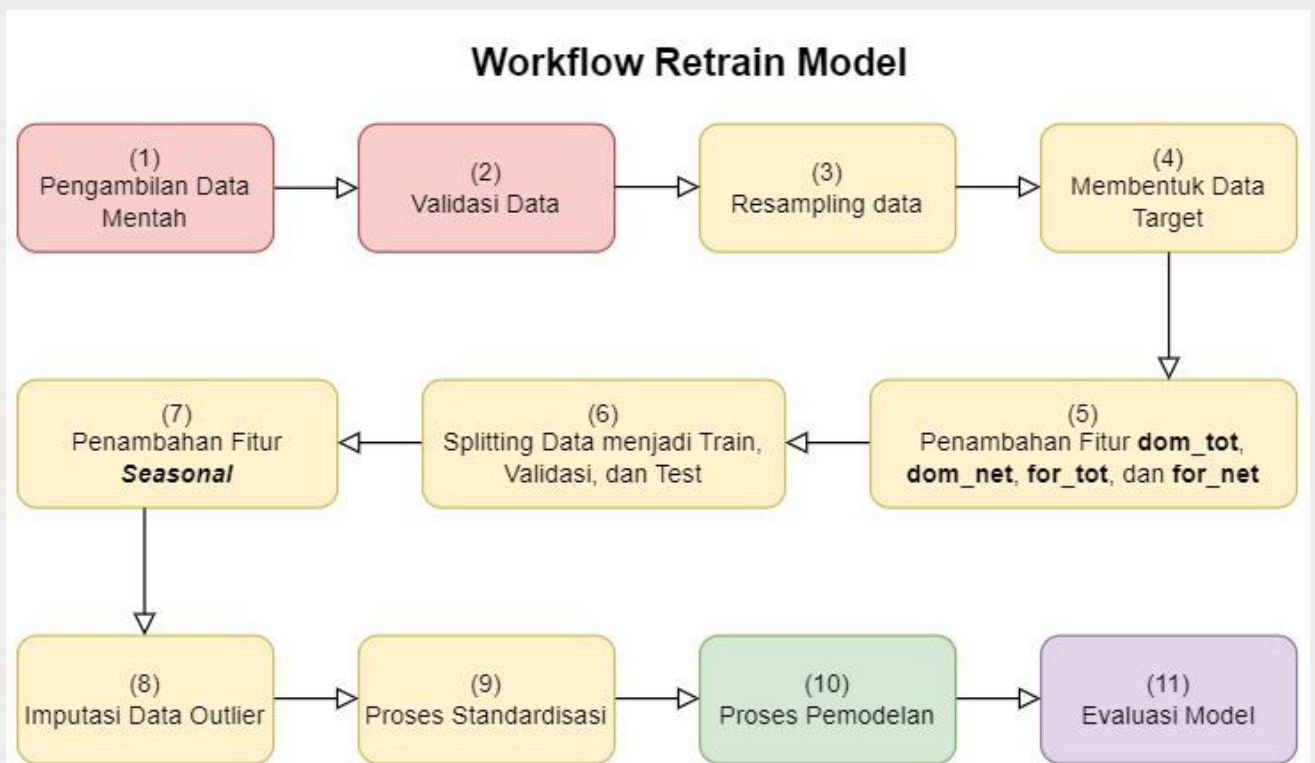


Gambar 16 : hasil prediksi pada data test

Dari *Gambar 15* dan *Gambar 16* dapat dilihat bahwa performa dari model dengan data yang tidak distasionerkan lebih baik daripada model saat data distasionerkan. Hasil prediksi mampu bergerak seiring dengan data yang sebenarnya meskipun masih belum terlalu akurat.

Dengan demikian, maka model best-fit yang akan digunakan untuk melakukan prediksi harga penutupan index LQ45 3 hari ke depan adalah model SARIMAX (1,0,3) (2,0,0) [20] tanpa melalui proses stasionerisasi.

Berikut pada *Gambar 17* merupakan diagram pipeline pemodelan.



Gambar 17 : workflow

5. KESIMPULAN

Model SARIMAX yang merupakan pengembangan dari model ARIMA/ARIMAX/SARIMA secara umum dapat digunakan untuk melakukan prediksi nilai indeks LQ45 3 hari ke depan.

Namun hasil prediksi dirasa masih kurang akurat.

Hal ini sebenarnya dapat diatasi dengan melakukan eksplorasi tiap-tiap fitur dengan lebih mendalam.

Bisa juga dengan melakukan hyperparameter tuning yang lebih variatif pada model.

Untuk penelitian lebih lanjut, dapat coba digunakan model machine learning LSTM (Long Short Term Memory network) dan/atau GRU (Gated Recurrent Unit) yang menurut beberapa sumber, dapat menghasilkan performa yang lebih baik ketimbang model SARIMAX.

Source code dan dokumentasi selengkapnya dapat diakses dpada repositori github :

https://github.com/ercainz/lq45_prediction

6. REFERENSI

6.1 DEKOMPOSISI SEASONAL & ADF TEST

- [statsmodels seasonal decompose](#)
- [How to Decompose Time Series Data into Trend and Seasonality](#) (by: Jason Brownlee)
- [Augmented Dickey Fuller Test \(ADF Test\) – Must Read Guide](#) (by: Selva Prabhakaran)

6.2 ARIMA, ARIMAX, SARIMA, SARIMAX

- [statsmodels SARIMAX](#)
- [pmdarima](#)
- [Time Series Forecasting with ARIMA , SARIMA and SARIMAX](#) (by: Brendan Artley)
- [How to Create an ARIMA Model for Time Series Forecasting in Python](#) (by: Jason Brownlee)
- [An Introduction to Time Series Analysis with ARIMA](#) (by: Taha Binhuraib)
- [Time Series Forecasting: ARIMA vs LSTM vs PROPHET](#) (by: Mauro Di Pietro)
- [ARIMA Model – Complete Guide to Time Series Forecasting in Python](#) (by: Selva Prabhakaran)

6.3 MODELLING DAN FORECASTING

- [Inflation Forecasting](#) (by: SwatiSethee)
- [Complete Guide To SARIMAX in Python for Time Series Modeling](#) (by: Yugesh Verma)