



ANALISIS FAKTOR-FAKTOR YANG MEMENGARUHI KUALITAS UDARA PERKOTAAN SERTA EVALUASI MODEL PREDIKSI BERBASIS KORELASI, REGRESI DATA PANEL, DAN DERET WAKTU

Presented by:

Kelompok
Josjis

Nama anggota:

Desi Nofitasari	24083010058
Diva Anggraeni	24083010065
Siti Rania Azaria	24083010072
Achmad Dany Gunawan	24083010075

STATISTIKA DESKRIPTIF

index	count	mean	min	max	std
PM2.5	175205	60,028	-30,550	145,072	19,993
PM10		89,970	-33,232	224,750	30,053
NO ₂		24,949	-23,087	68,019	9,964
SO ₂		10,002	-6,575	28,328	3,995
CO		599	-357	1,499	199
O ₃		30,001	-11,472	75,843	10,002
Temp_C		27,005	3,083	48,162	4,992
Humidity_%		64,957	-3,501	130,720	15,001
Wind_Speed_m_ps		2,503	-1,857	7,156	1,003
Pressure_hPa		1,009,995	987,584	1,033,106	4,996
Rain_mm		94	0	10	238
AQI_Target		43,499	-8,410	89,239	10,627

Berdasarkan statistika deskriptif, PM2.5 dan PM10 memiliki standar deviasi lebih besar dibanding variabel lain, menunjukkan variasi data tinggi serta indikasi nilai ekstrem. Polutan gas umumnya stabil, kecuali O₃ yang memiliki variasi lebih besar. Variabel meteorologis relatif stabil, sedangkan rata-rata AQI berada pada kategori baik dengan variasi sedang.

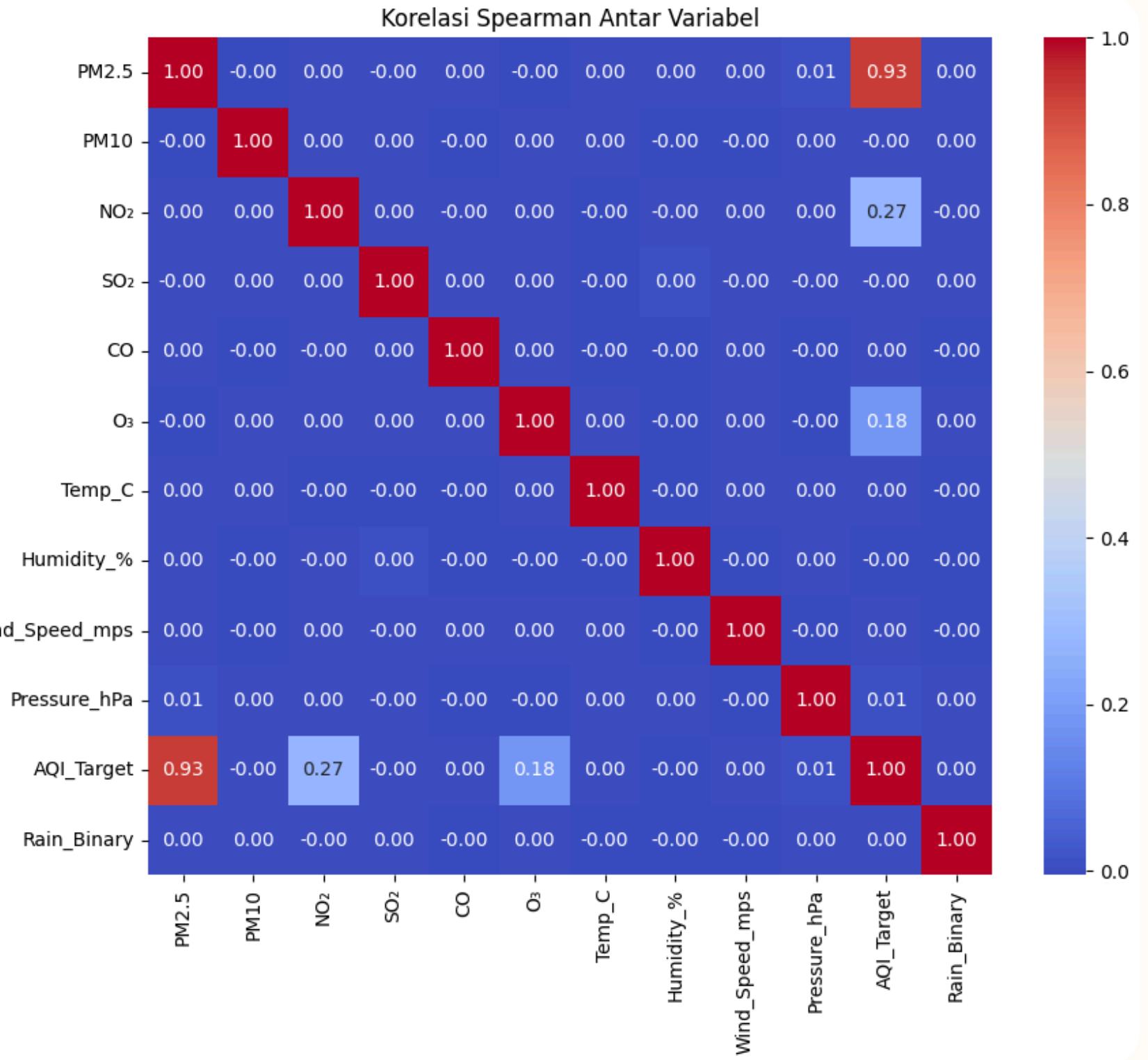
PEMILIHAN METODE KORELASI

Variabel	shapiro_pvalue	summary
PM2.5	67,081	Normal
PM10	65,060	Normal
NO ₂	79,390	Normal
SO ₂	98,487	Normal
CO	96,245	Normal
O ₃	82,459	Normal
Temp_C	91,387	Normal
Humidity_%	92,174	Normal
Wind_Speed_mps	98,587	Normal
Pressure_hPa	84,599	Normal
AQI_Target	83,260	Normal
Rain_Binary	0	Not Normal

Uji normalitas awal dilakukan menggunakan Shapiro-Wilk karena sensitif dalam mendeteksi penyimpangan distribusi pada variabel numerik kontinu. Hasil menunjukkan mayoritas variabel numerik mendekati normal, sementara variabel biner tidak normal. Dengan demikian, korelasi spearman dipilih karena tidak mensyaratkan normalitas dan lebih andal untuk dataset besar.

UJI KORELASI

Hasil korelasi Spearman menunjukkan PM2.5 memiliki korelasi sangat kuat dan positif (0,93) terhadap AQI_Target, sehingga kenaikan PM2.5 meningkatkan AQI. Korelasi sedang terdapat pada NO₂ (0,27) terhadap AQI, sedangkan polutan lain dan variabel meteorologi memiliki korelasi lemah hingga tidak berpengaruh, menunjukkan dominan PM2.5 terhadap kualitas udara di wilayah pengamatan.



REGRESI DATA PANEL

UJI CHOW & UJI HAUSMAN

	Cross-section F	P-Value
Uji Chow	1.8189	0.1220
Uji Hausman	18.29756	0.00108

Uji Chow menunjukkan p-value > 0.05 sehingga model Common Effect tidak dapat ditolak. Namun, uji Hausman menghasilkan p-value < 0.05 yang mengindikasikan bahwa Fixed Effect Model lebih tepat digunakan dibandingkan Random Effect Model dalam analisis regresi data panel.

REGRESI DATA PANEL

MODEL FIXED EFFECT

Variabel	Kode	Coefficients
Humidity_%	X ₁	-0.00007228
PM2.5	X ₂	0.5001
O ₃	X ₃	0.2006
PRESSURE_HPA	X ₄	0.0008
CO	X ₅	-0.0135

Pada model Fixed Effect, PM2.5 dan O₃ memiliki koefisien positif terbesar sehingga berkontribusi besar terhadap peningkatan AQI. Humidity dan CO₃ menunjukkan pengaruh negatif, sedangkan tekanan udara berpengaruh positif, namun relatif kecil terhadap variasi AQI

REGRESI DATA PANEL

UJI ASUMSI KLASIK

Jenis Uji	Metode/statistik Uji	Nilai Statistik	P-Value	Keputusan
Normalitas	Shapiro- Wilk	0.99998	0.7950	Gagal Tolak H_0
Heteroskedastisitas	Breusch-Pagan	2.4895	0.7781	Gagal Tolak H_0
Autokorelasi	Durbin-Watson	2.0034	...	Tidak terjadi autokorelasi
Multikolinearitas	Variance Inflation Factor (VIF)	VIF maks 1.000038	...	Tidak terjadi multikolinearitas

Hasil uji asumsi klasik menunjukkan residual berdistribusi normal dan tidak mengalami heteroskedastisitas. nilai Durbin-Watson mendekati 2 menandakan tidak terjadi autokorelasi, serta nilai VIF yang rendah menunjukkan tidak adanya multikolinearitas. Model regresi panel memenuhi asumsi klasik.

REGRESI DATA PANEL

UJI HIPOTESIS REGRESI

Uji F-statistik

	F-test	p-value
F-statistic	407.800	0.0000

Uji F-statistik menghasilkan p-value < 0.05 yang menunjukkan model regresi panel signifikan secara simultan. Uji t-statistik menunjukkan bahwa PM2.5 dan O₃ berpengaruh signifikan terhadap AQI, sementara Humidity, Pressure, dan CO tidak berpengaruh signifikan secara parsial.

Uji t-statistik

Variabel	Kode	p-value
Humidity_%	X ₁	0.8793
PM2.5	X ₂	0.0000
O ₃	X ₃	0.0000
Pressure_hPa	X ₄	0.5656
CO	X ₅	0.7066

REGRESI DATA PANEL

UJI KOEFISIEN DETERMINASI

Model Regresi	R ²	Keterangan
Fixed Effect	0.9209	Model mampu menjelaskan 92% variasi AQI

Nilai R² Sebesar 0.9209 menunjukkan bahwa model regresi data panel fixed effect memiliki kemampuan penjelasan yang sangat kuat. Serta dapat dijelaskan oleh variabel independen dalam model

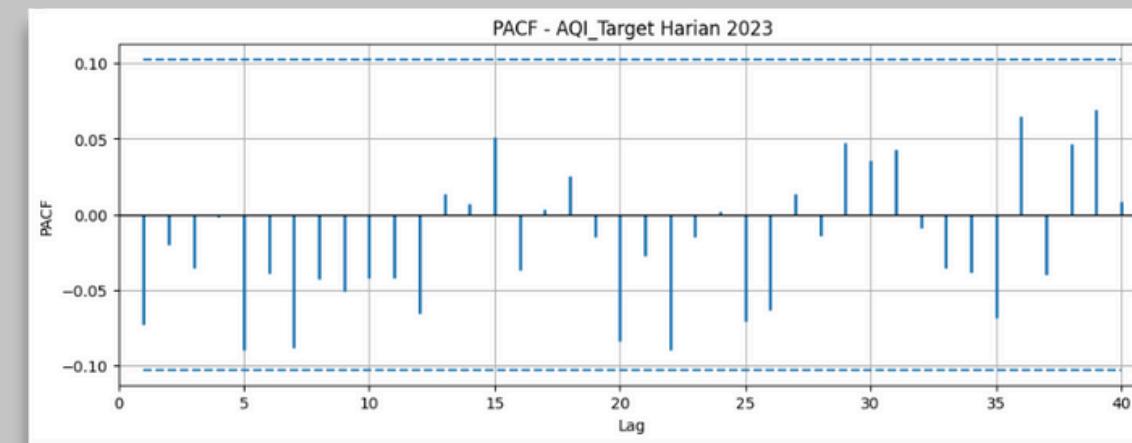
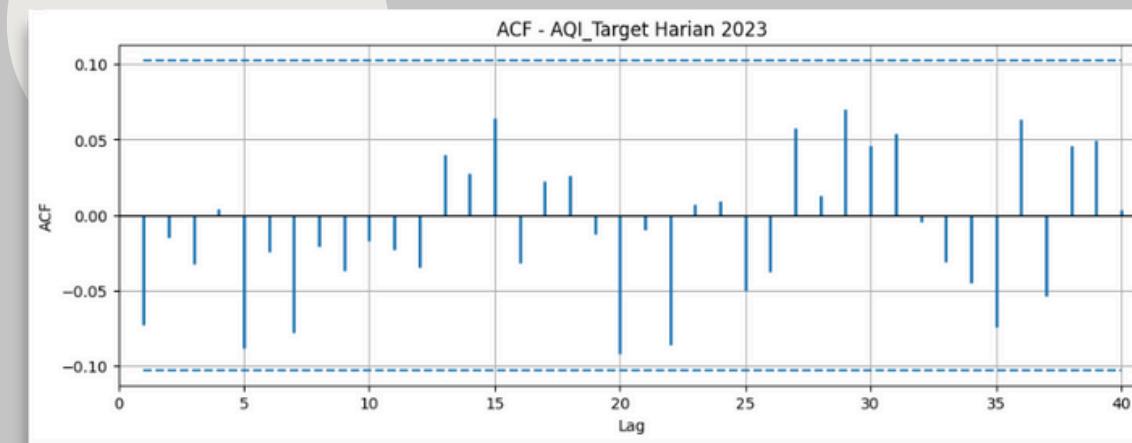
ANALISIS DERET WAKTU

UJI STASIONERITAS

ADF	-20.517095071253582
p-value	0,0

Nilai ADF sebesar -20,517 dengan p-value 0,0 menunjukkan bahwa H_0 ditolak. Dengan begitu, data telah stasioner dan tidak diperlukan *differencing* sebelum pemodelan dilakukan.

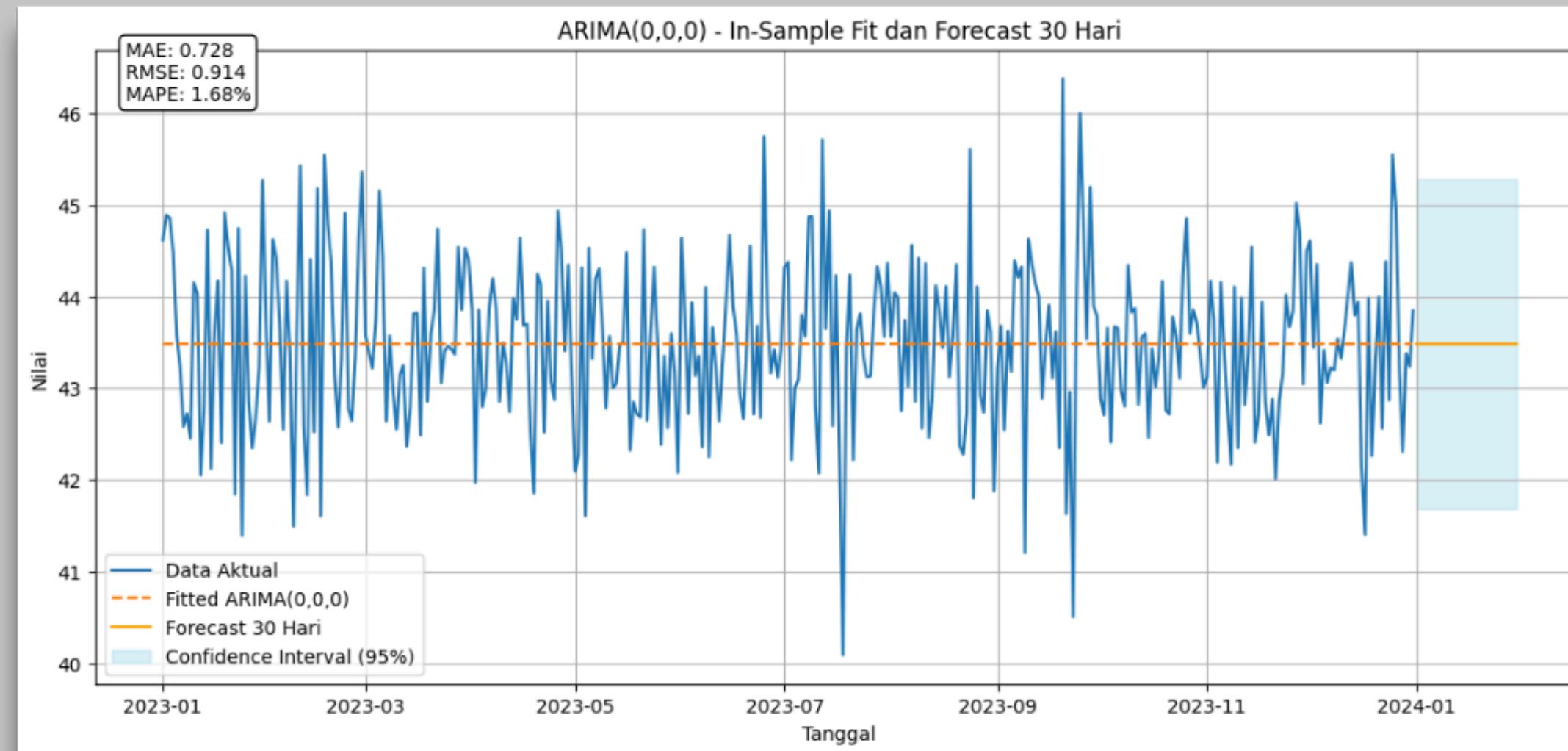
WHITE NOISE



Plot ACF dan PACF menunjukkan seluruh spike berada di dalam batas kepercayaan. Hal tersebut menandakan tidak adanya autokorelasi yang signifikan, sehingga data cenderung *white noise*.

ANALISIS DERET WAKTU

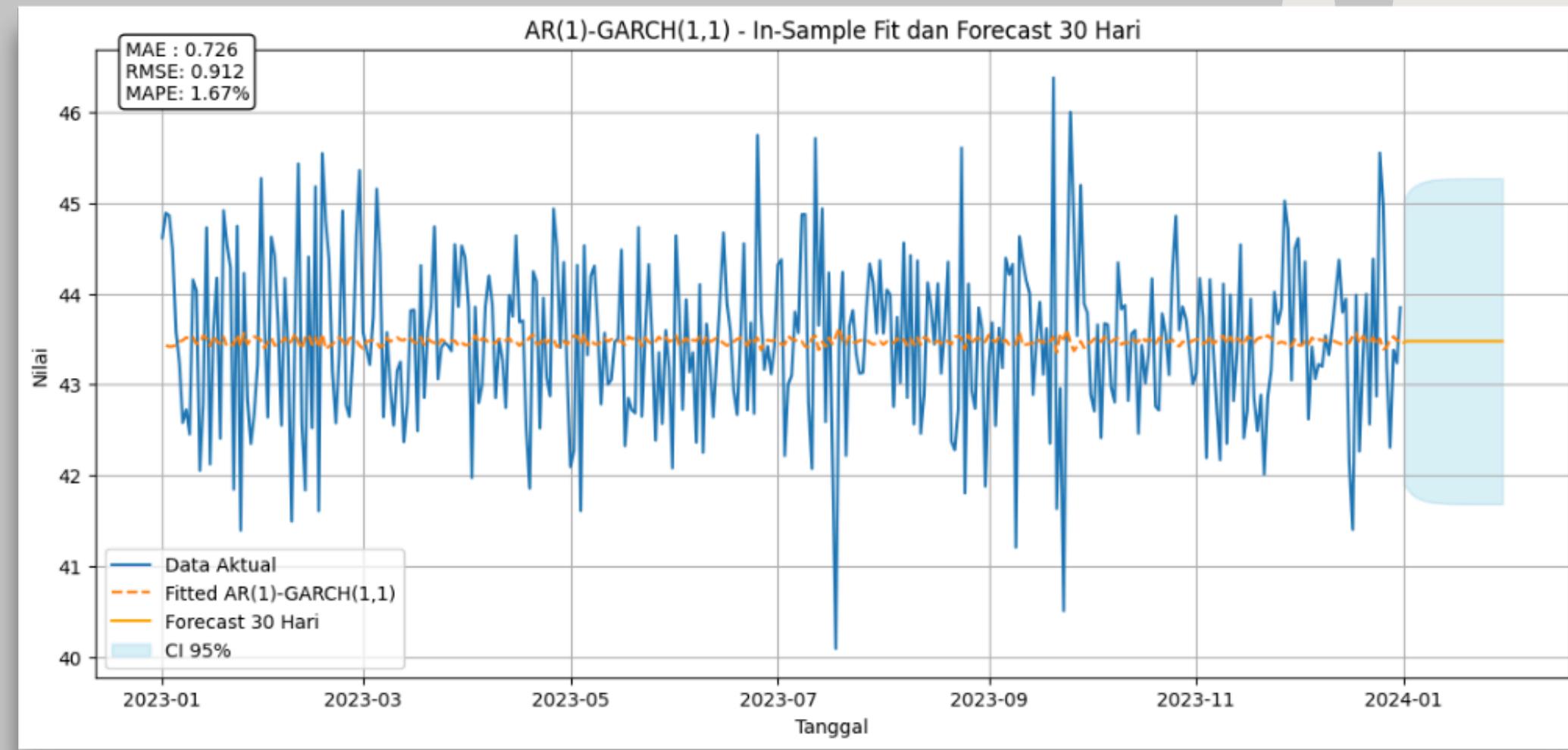
MODEL ARIMA



Model ARIMA(0,0,0) merepresentasikan nilai rata-rata data, sehingga garis *fitted* dan *forecast* terlihat mendatar. Oleh karena itu model ini kurang mampu menangkap fluktuasi data aktual di sekitar rata-rata.

ANALISIS DERET WAKTU

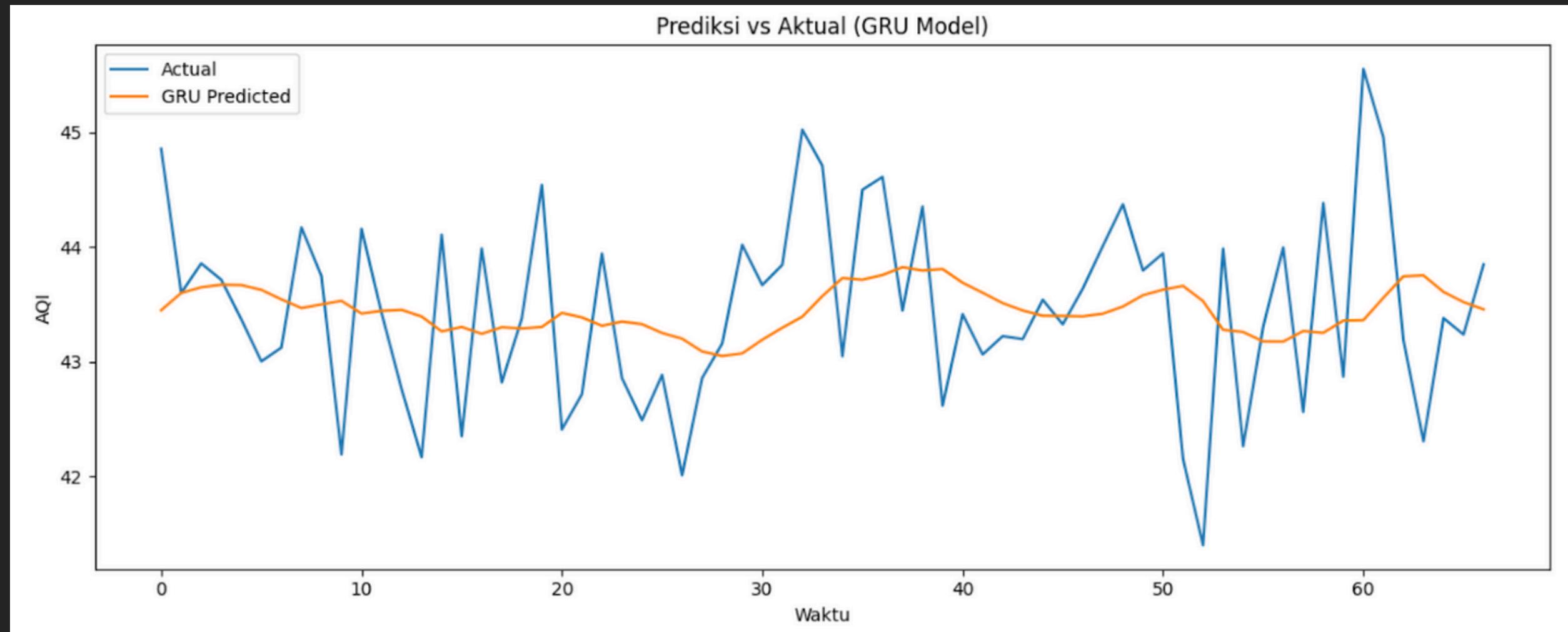
MODEL AR-GARCH



Hasil model AR(1)-GARCH(1,1) menunjukkan garis *fitted* yang sedikit berfluktuasi mengikuti dinamika data. Hasil *forecast* tetap berada di sekitar rata-rata, namun interval kepercayaan menyesuaikan perubahan volatilitas sehingga menggambarkan ketidakpastian data secara lebih realistik.

MACHINE LEARNING FORECAST

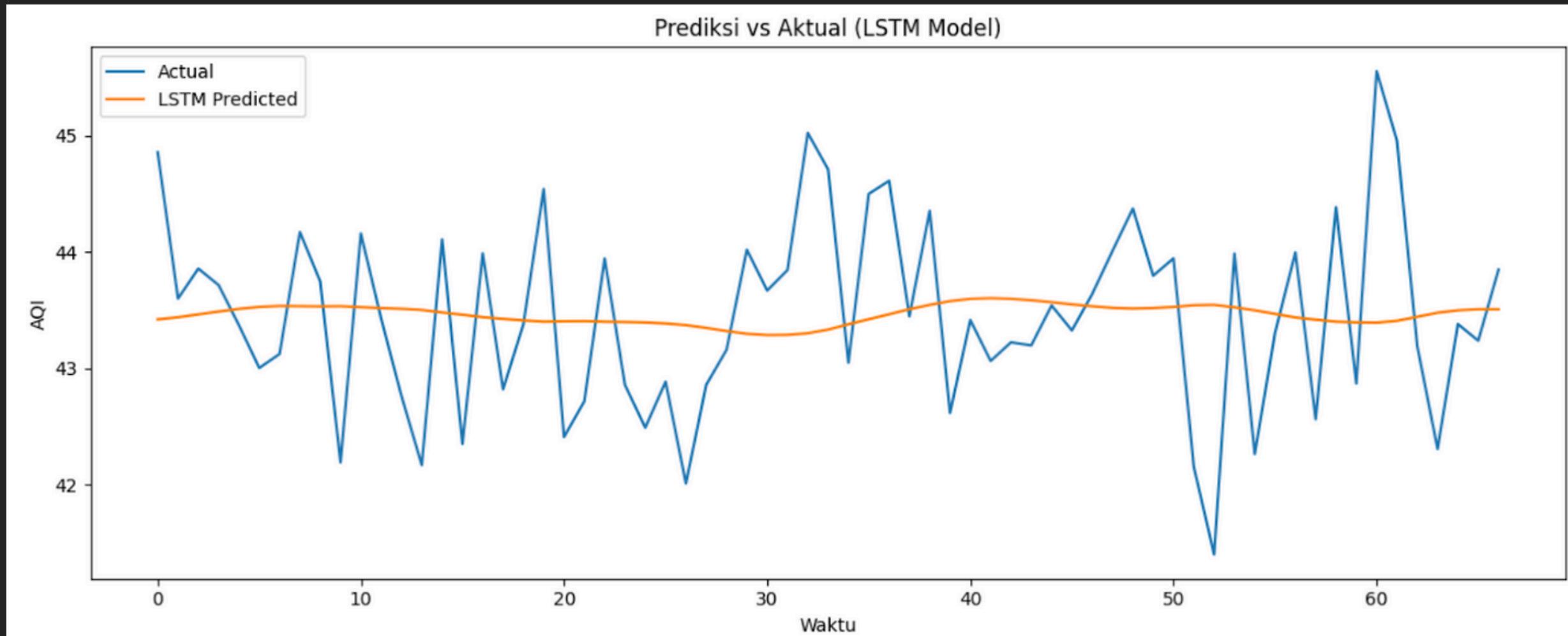
GRU



Hasil plot menunjukkan garis prediksi yang sedikit mengikuti perubahan fluktuasi data aktual. Jarak vertikal antara prediksi dan nilai aktual yang dihasilkan relatif kecil. Hal ini ditandai pada garis prediksi pada titik waktu 30 hingga 45 yang bergerak dengan arah yang sama sesuai garis data aktual.

MACHINE LEARNING FORECAST

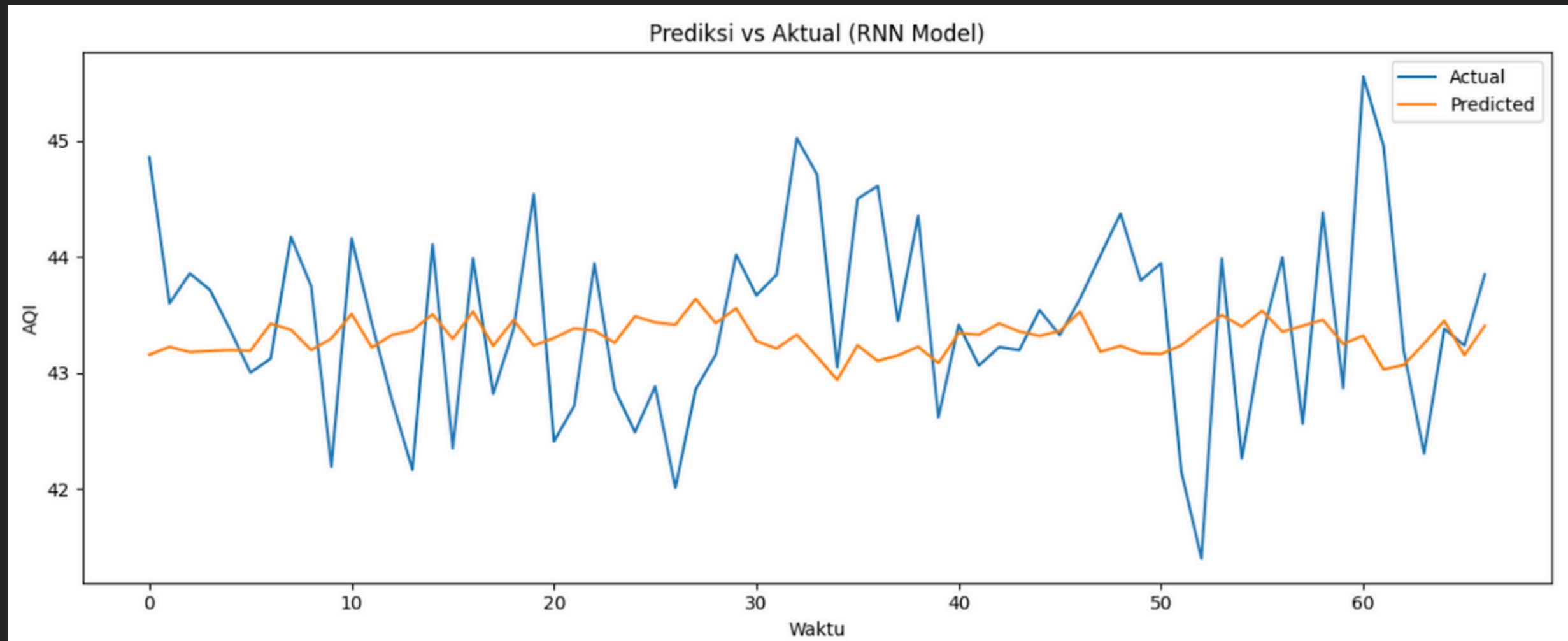
LSTM



Model hasil LSTM menunjukkan bahwa prediksi tidak mampu menangkap fluktuasi sebaik GRU. Hal ini dikarenakan model LSTM terlalu banyak melakukan smoothing untuk mempertahankan garis prediksi di nilai rata-rata. Berdasarkan gambar, nilai rata-rata AQI yang dipertahankan berada pada 43,5.

MACHINE LEARNING FORECAST

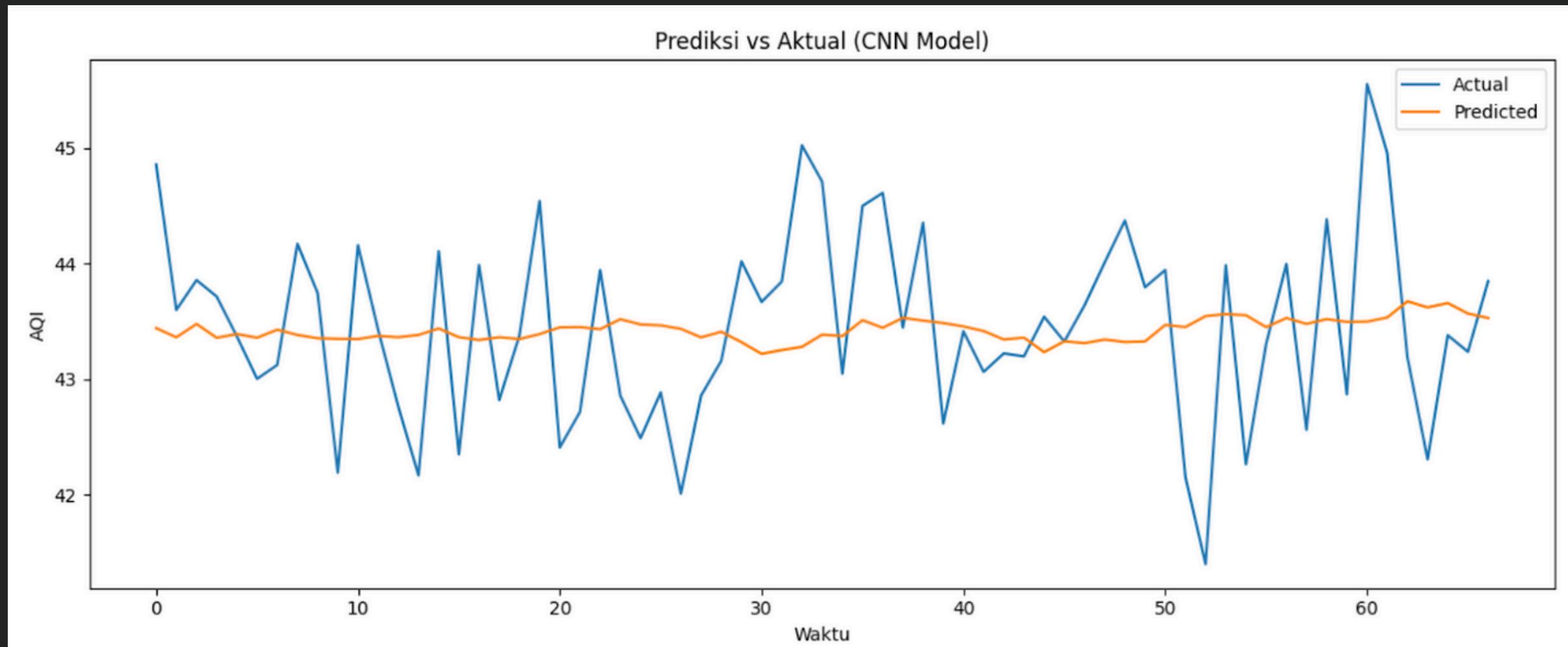
RNN



Model RNN menunjukkan prediksi yang secara umum mampu mengikuti pola fluktuasi aktual. Namun, terlihat selisih yang cukup jelas pada perubahan nilai (titik waktu 52). Hal ini dikarenakan prediksi RNN hanya dipengaruhi oleh $t-1$ dan $t-2$ yang memiliki AQI lebih tinggi, sehingga RNN kurang merespon penurunan esktrem di titik 52.

MACHINE LEARNING FORECAST

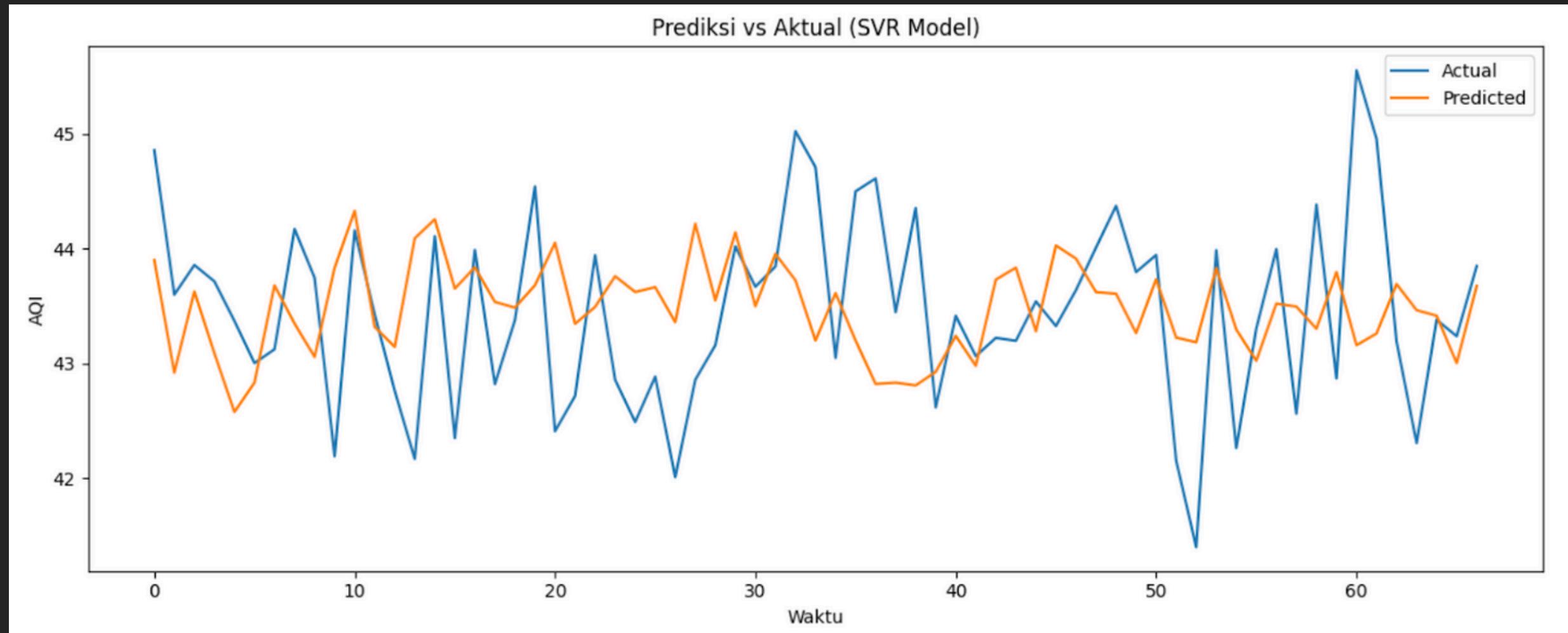
CNN



Plot CNN menunjukkan kurva prediksi yang lebih halus daripada RNN. Namun, kurang mampu menangkap fluktuasi ekstrem sebaik model RNN. Hal ini dikarenakan CNN memproses nilai secara simultan, sehingga fluktuasi tajam pada satu titik tertentu cenderung tidak terlihat.

MACHINE LEARNING FORECAST

SVR



Dibandingkan dengan beberapa model sebelumnya, SVR memprediksi data dengan jarak paling jauh dari data aktual. Terlihat pada plot terdapat deviasi yang menyimpang (titik waktu 60). Selain itu, di titik awal terlihat bahwa model SVR terlalu mengikuti fluktuasi tajam pada jangka pendek (bukan menangkap sebagai pola).

EVALUASI MODEL

ARIMA, AR-GARCH, MACHINE LEARNING FORECAST

	RMSE	MAE	MAPE
ARIMA	0,914	0,728	1,68%
AR-GARCH	0,912	0,726	1,67%
GRU	0,13353	0,10900	23,11%
LSTM	0,13418	0,10849	23,11%
RNN	0,13814	0,11106	22,64%
CNN	0,13393	0,10973	23,24%
SVR	0,14620	0,11730	24,95%

Model ARIMA & AR-GARCH memiliki nilai metrik evaluasi yang lebih besar dibandingkan model dengan pendekatan *machine learning forecast*. Hal ini menunjukkan bahwa pendekatan dengan *machine learning forecast* lebih efektif.

Hasil evaluasi menunjukkan bahwa model GRU merupakan model terbaik yang ditandai dengan nilai RMSE sebesar 0,13353.

TERIMA KASIH