NANYANG TECHNOLOGICAL UNIVERSITY
AI6123-TIME SERIES ANALYSIS
Assignment1 Individual
NTAMBARA Etienne G2304253K
February 29, 2024

NTAMBARA Etienne G2304253K [1 2 3] February 29, 2024

1. Intoduction

The wwwusage time series data consist of the number of users connected to the internet through a server. The data are collected at a time interval of one minute and there are 100 observations. I have conducted different fitted ARIMA models using R. The fitted models and diagnostic checking, AIC, BIC, and HQC results are presented in this report.

2. Results

2.1. Data Loading

The first step in order to work on our wwwusage dataset in to load data, this is the result of loaded data:

```
Loaded Dataset
----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
  [1]  88  84  85  85  84  85  83  85  88  89  91  99 104 112 126 138 146 151
 [19] 150 148 147 149 143 132 131 139 147 150 148 145 140 134 131 131 129 126
 [37] 126 132 137 140 142 150 159 167 170 171 172 172 174 175 172 172 174 174
 [55] 169 165 156 142 131 121 112 104 102  99  99  95  88  84  84  87  89  88
 [73]  85  86  89  91  91  94 101 110 121 135 145 149 156 165 171 175 177 182
 [91] 193 204 208 210 215 222 228 226 222 220
```

Figure 1. Loaded dataset.

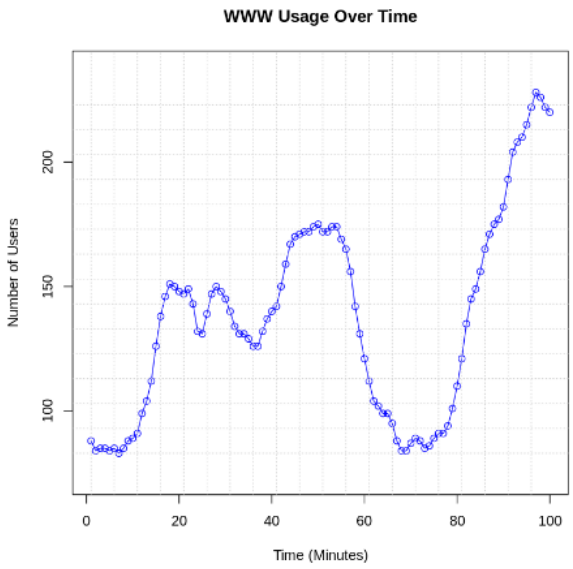

Figure 2. Original Plot of dataset.

2.2. Check Stationary or Nan-Stationary using Plot

To know that our time series is stationary we need to testify if its mean is constant over time and Covariance is not dependent on time lag and check if constant autocorrelation structure correlated with its past values should not change over time. Otherwise, our time series is non-stationary. Now let's analyse the below plot from the original dataset:

Above figure 2 shows that our time series is not stationary because between time 20-40 and 60-80 the variance is very high that's means the change in variance over time. Also, the mean is not constant because of upward trending.

2.3. ACF and PACF Plot for Original Dataset

To determine the order of AR and MA terms in ARIMA model, which is important for modeling and forecasting the time series data we need to plot ACF and PACF. plotted results shown below:

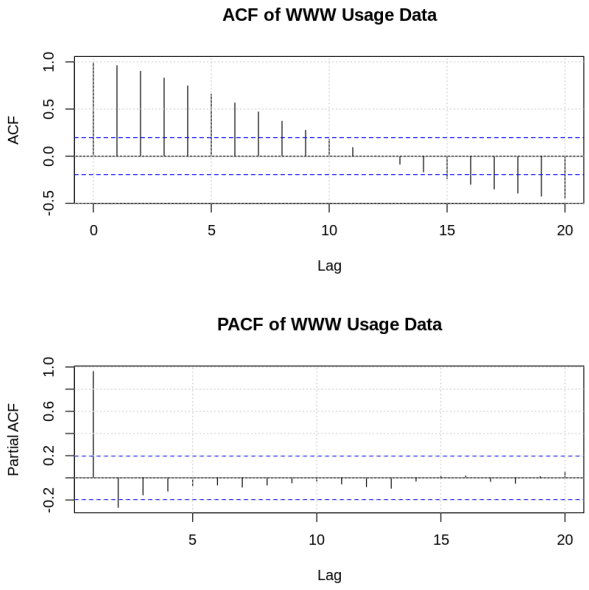

Figure 3. ACF and PACF Orginal data plot.

Our ACF from figure 3 .....and PCF 3 has cut off at lag of 2.

3. Diffencing

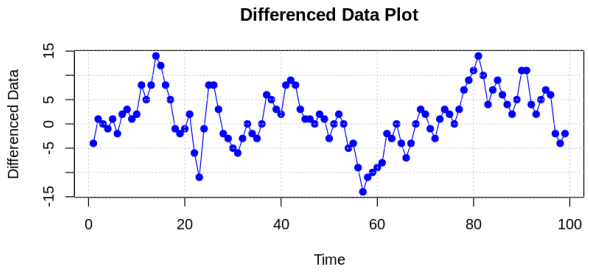

Figure 4. Differencing Data.

The above figure 4 of differencing shows that our time series's data is now stationary.

3.1. ACF AND PACF after Differencing

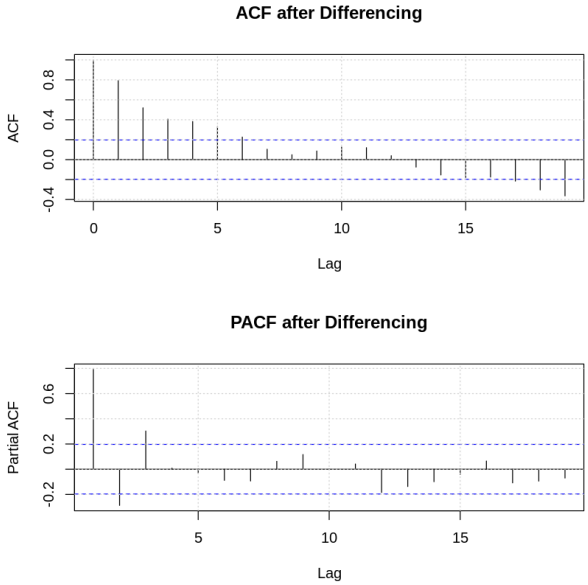

Figure 5. ACF AND PACF after Differencing

3.2. ARIMA(1,1,1) Dignostic one time differencing

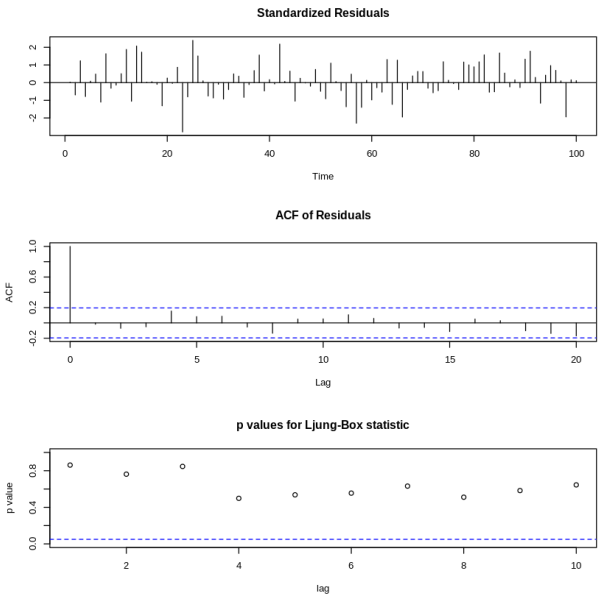

Figure 6. ARIMA(1,1,1) Dignostic

3.3. ARIMA(3,1,0) Dignostic one time differencing

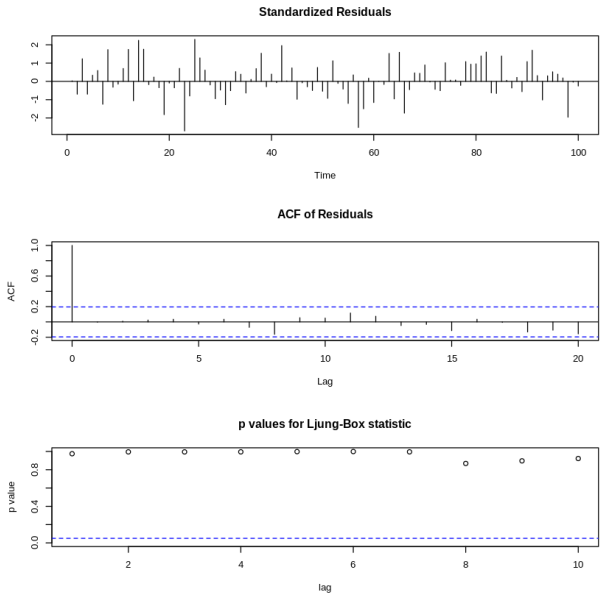

Figure 7. ARIMA(3,1,0) Dignostic

3.4. ARIMA MODEL DIAGNOSTICS AIC two time differencing

ARIMA(5,2,5)

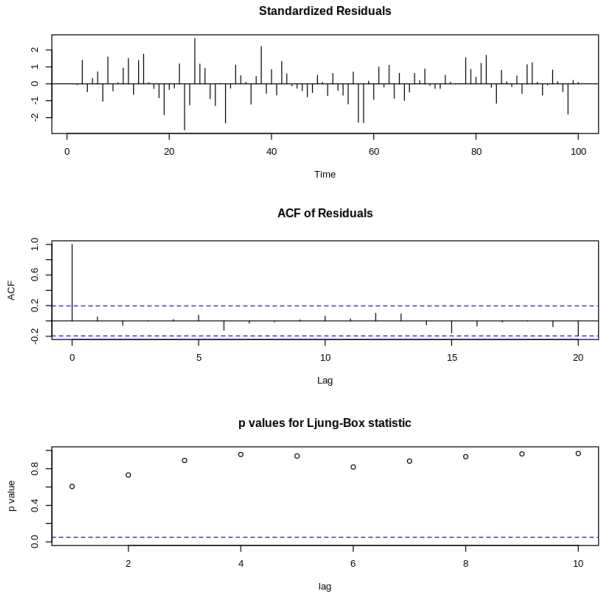

Figure 8. ARIMA(5,2,5)

3.5. ARIMA MODEL DIAGNOSTICS BIC/HQC two time differencing

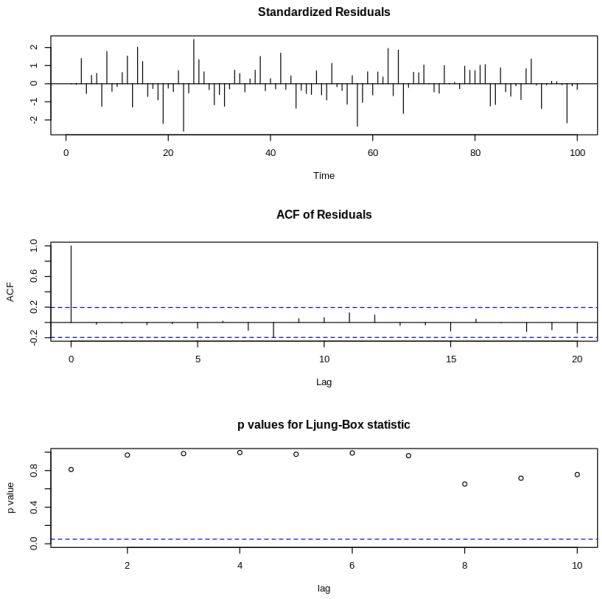

Figure 9. ARIMA(2,2,0) Diagnostic

3.6. AIC,BIC and HQC Fitted Values and forecasted values analysis

Table 1. AIC, BIC, and HQC of the fitted models

Model	ARIMA	AIC	BIC	HQC
1	(3,1,0)	511.994	522.3745	513.1571
2	(1,1,1)	514.2995	522.0848	514.4082
3	(2,2,0)	511.4645	519.2194	511.5733
4	(5,2,5)	509.8191	538.2538	518.3627

.....GPT4 FOR EXPLANATION.....

3.7. Plot all above fitted models

table 1 show the results of all fitted models, now let visualise them on below plots.

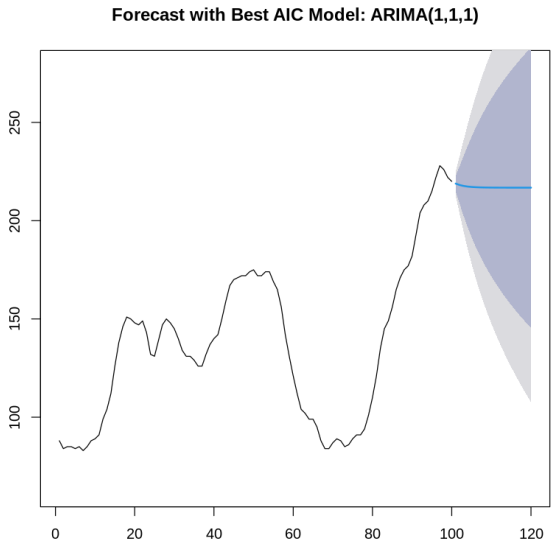

Figure 10. Forecast fitted for ARIMA(1,1,1)

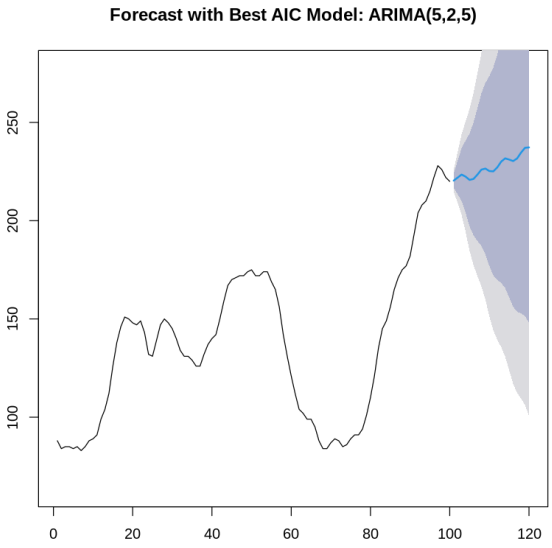

Figure 12. Forecast fitted for ARIMA(5,2,5)

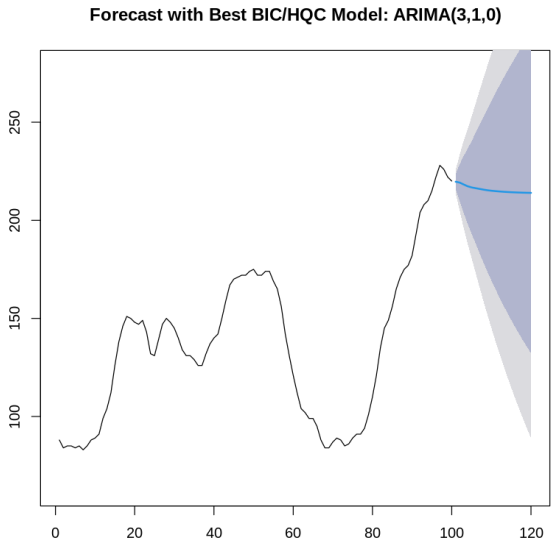

Figure 11. Forecast fitted for ARIMA(3,1,0)

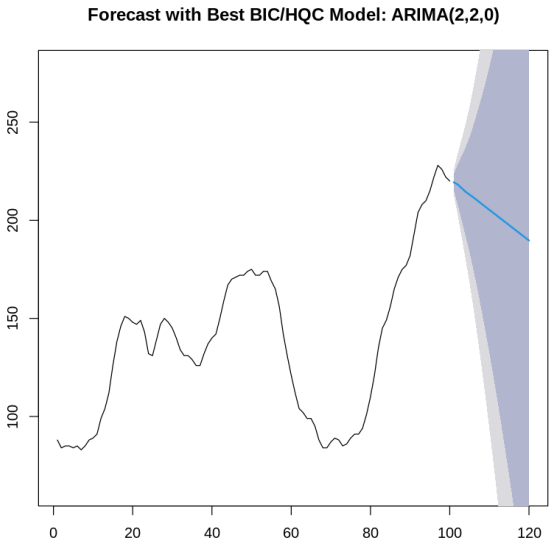

Figure 13. Forecast fitted for ARIMA(2,0,2)

3.8. Model Accuracy's Report

Accuracy Assessment for Model ARIMA(1,1,1):

A matrix: 2 × 8 of type dbl

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0.3035616	3.113754	2.405275	0.2805566	1.917463	0.5315228	-0.01715517	NA
Test set	-2.5855588	11.351388	9.265086	-1.4666403	4.437330	2.0474185	0.64113200	1.477598

Accuracy Assessment for Model ARIMA(3,1,0):

A matrix: 2 × 8 of type dbl

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0.230588	3.044632	2.367157	0.2748377	1.890528	0.5230995	-0.003095065	NA
Test set	-2.168862	12.071916	10.086919	-1.2910728	4.816833	2.2290289	0.659781410	1.616196

Figure 14. Forecast fitted for ARIMA(1,1,1) and ARIMA(3,1,0)

Accuracy Assessment for Model ARIMA(5,2,5):

A matrix: 2 × 8 of type dbl

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0.1330769	2.714929	2.097323	0.2171532	1.640865	0.463471	0.05077024	NA
Test set	-8.3453303	12.521428	9.401544	-4.1107056	4.575738	2.077573	0.64418970	1.750413

Accuracy Assessment for Model ARIMA(2,2,0):

A matrix: 2 × 8 of type dbl

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0.02797758	3.150308	2.511921	0.2062350	1.994727	0.5550897	-0.0235521	NA
Test set	2.39255695	14.750798	13.118737	0.7787425	6.155561	2.8990065	0.6808541	2.186538

Figure 15. Forecast fitted for ARIMA(5,2,5) and ARIMA(2,2,0)

3.9. Residual Results Analysis for each Model

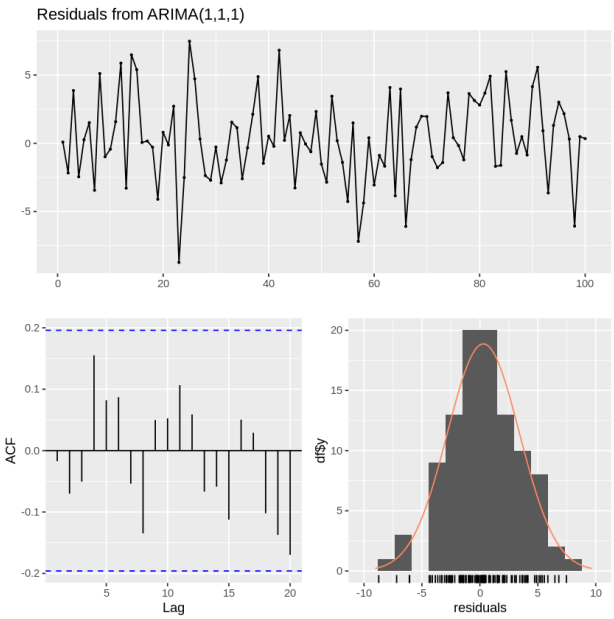

Figure 16. Residual ARIMA(1,1,1)

Figure 16, The ARIMA(1,1,1) model's diagnostic plots demonstrate a strong match; residuals show up as white noise with no discernible patterns or systematic structure. The Ljung-Box test, which yields a high p-value of 0.4499,lags of 10, and Q* = 7.8338, indicates that there is no significant autocorrelation in the residuals, as suggested by the ACF plot. The model's assumptions are supported by the residual distribution, which closely approaches a normal distribution.

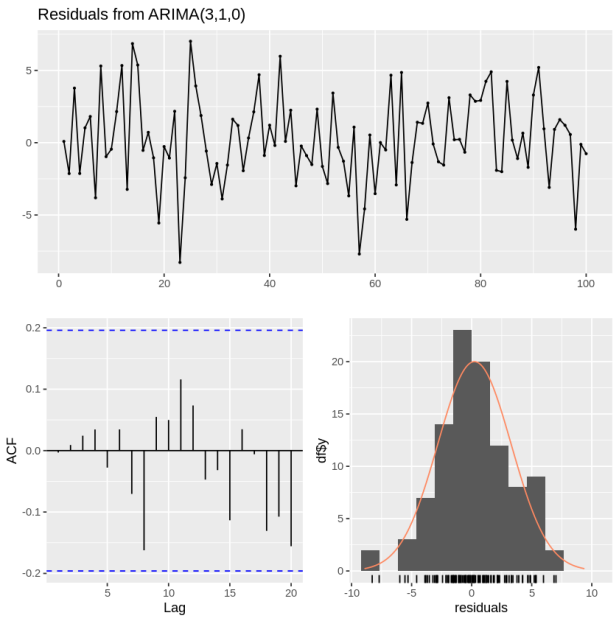

Figure 17. Residual ARIMA(3,1,0)

Figure 17,The residuals for an ARIMA(3,1,0) model are randomly distributed around zero, with no discernible trend or seasonality, according to the diagnostic plots. The autocorrelations are within the confidence bounds, as indicated by the ACF plot, indicating a well-fitting model. The Ljung-Box test findings validate the validity of the model by confirming that there is no substantial autocorrelation in the residuals, with a high p-value of 0.7218 and total lags of 10, Q* = 4.4913.

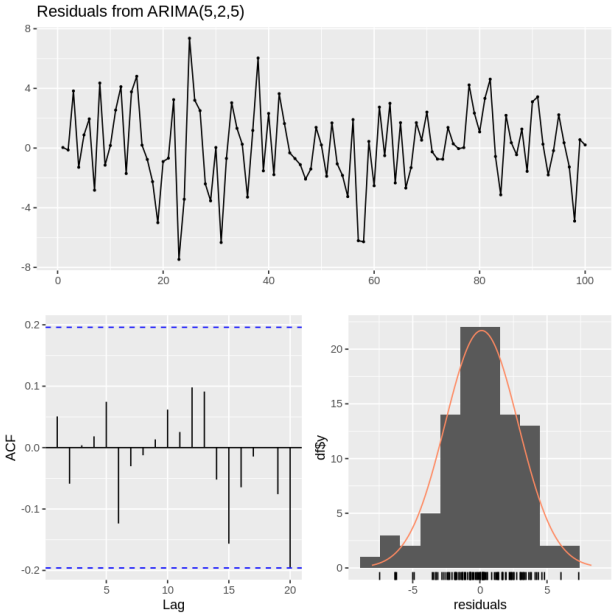

Figure 18. Residual ARIMA(5,2,5)

The diagnostic figure 18 for the ARIMA(5,2,5) model display a histogram that approximates a normal distribution, randomly behaved residuals, and ACF values within confidence bounds indicating no autocorrelation. With a p-value of 0.1297, the Ljung-Box test shows that there is no discernible autocorrelation in the residuals. These diagnostics collectively imply that the ARIMA(5,2,5) model provides a sufficient fit to the data.

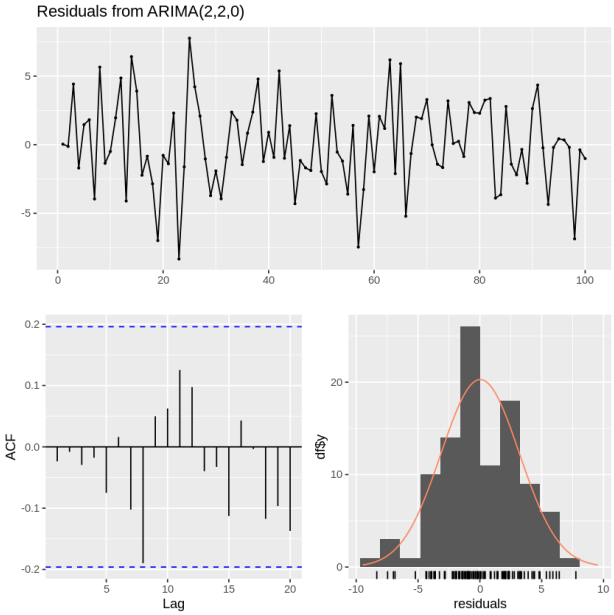

Figure 19. Residual ARIMA(2,2,0)

Figure 19 The ARIMA(2,2,0) model's residuals randomly oscillate around zero, demonstrating how well the model fits the data. Since all of the spikes in the ACF plot are contained within the confidence intervals, there is no discernible autocorrelation. The appropriateness of the model is confirmed by the Ljung-Box test, which yielded a p-value of 0.5717,Q* = 6.6784, indicating the absence of significant autocorrelation in the residuals.

Acknowledgement

Codes

References

[1]Nanyang Technological University (NTU), Singapore [2]School of Computer Science and Engineering (SCSE) [3]Master of Science in Artificial Intelligence. Correspondence to: Ntambara Etienne <ntam0001@e.ntu.edu.sg>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.