**Usecase : Peak Detection**

**CMX1\_AL\_CP1.csv :-**

# of time points – 433666

Step 1 – Global Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| CMX1\_AL\_CP1.csv | 0.9996 | 0.8577 | 0.0172 | 0.3800 |

Step 2 – Peak Detection

Autoencoder # of Peaks – 95150 [Window\_Size = 30, Step = 5, Latent\_Space = 16]

Z-Score # of Peaks – 278596 [Lag = 5000, Influence=0.5, Threshold =0.5xMean]

Step 3 – Statistical Analysis

For Autoencoder

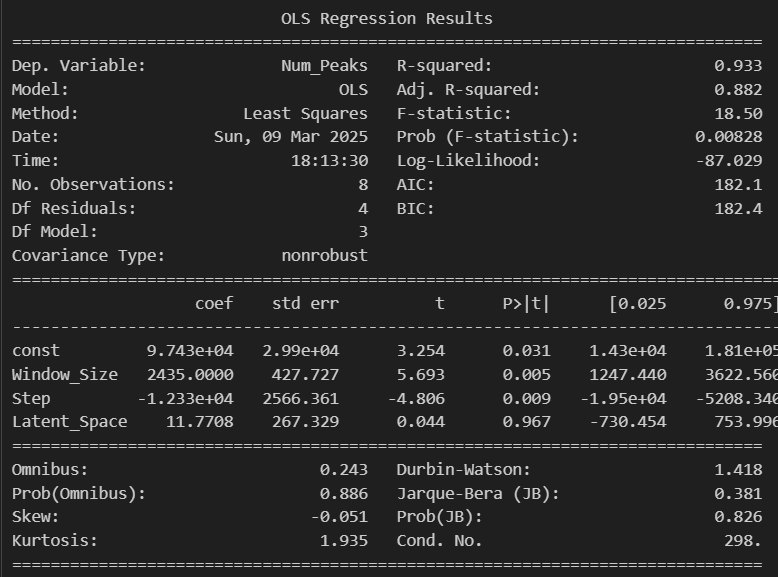
* 1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Window\_Size) | 1.0673E+10 | 1 | 55.7851141 | 0.08473169 |
| C(Step) | 7606007785 | 1 | 39.7561806 | 0.10013262 |
| C(Latent\_Space) | 638450 | 1 | 0.00333714 | 0.96326457 |
| C(Window\_Size):C(Step) | 918961321 | 1 | 4.80335983 | 0.27251185 |
| C(Window\_Size):C(Latent\_Space) | 858050 | 1 | 0.00448498 | 0.95742915 |
| C(Step):C(Latent\_Space) | 206105904 | 1 | 1.07730413 | 0.48815178 |
| Residual | 191316361 | 1 |  |  |

* 1. Tukey’s Test

Not Applicable since no significant features were found.

* 1. Multinomial Regression



For Z-Score

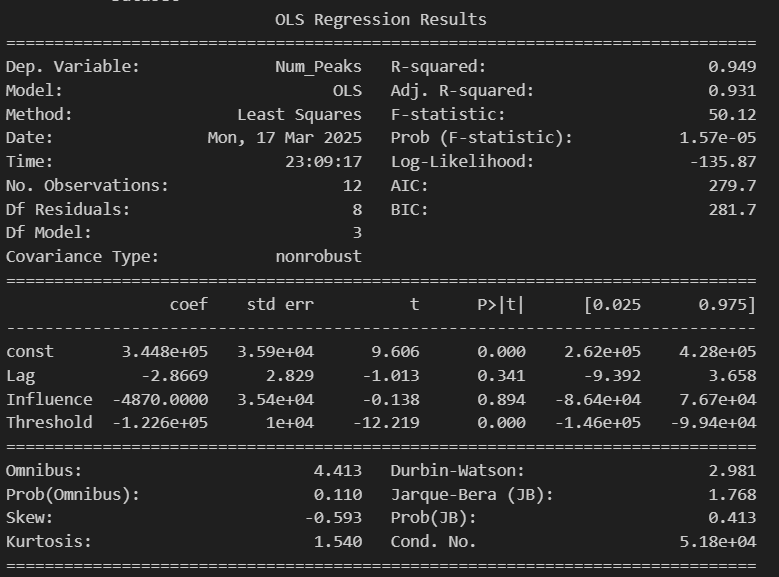
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Lag) | 616419336 | 1 | 926.781106 | 0.00107726 |
| C(Influence) | 11384112 | 1 | 17.1159133 | 0.05375736 |
| C(Threshold) | 9.4372E+10 | 2 | 70943.3753 | 1.41E-05 |
| C(Lag):C(Influence) | 4076336.33 | 1 | 6.12873619 | 0.13169171 |
| C(Lag):C(Threshold) | 70932506.2 | 2 | 53.3232028 | 0.01840834 |
| C(Influence):C(Threshold) | 2844454.5 | 2 | 2.13830629 | 0.31864321 |
| Residual | 1330237.17 | 2 |  |  |

1. Tukey’s Test

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group1 | group2 | meandiff | p-adj | lower | upper | reject | Factor |
| 5000 | 10000 | -14334.3333 | 0.8036 | -139363.225 | 110694.558 | False | Lag |
| 0.56543053 | 1.13086105 | -115225 | 0 | -132722.893 | -97727.1075 | True | Threshold |
| 0.56543053 | 2.2617221 | -217085.75 | 0 | -234583.643 | -199587.858 | True | Threshold |
| 1.13086105 | 2.2617221 | -101860.75 | 0 | -119358.643 | -84362.8575 | True | Threshold |
| 5000 | 10000 | -14334.3333 | 0.8036 | -139363.225 | 110694.558 | False | Lag |

1. Multinomial Regression



Step 4 – Classification (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | train\_accuracy | test\_accuracy | train\_f1 | test\_f1 |
| Data1\_Classes.csv RF AE | 1 | 0.78775441 | 1 | 0.69794592 |
| Data1\_Classes.csv RF Z-Score | 1 | 0.53604298 | 1 | 0.5487783 |

Step 5 – Regression (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| Peak\_LSTM\_RF\_CMX1\_AL\_CP1.csv | 0.99937928 | 0.92910308 | 0.0334554 | 0.45679939 |
| NonPeak\_LSTM\_RF\_CMX1\_AL\_CP1.csv | 0.9793477 | 0.93499536 | 0.06387419 | 0.25153225 |
| Peak Aware Model - AE |  |  |  |  |
| Peak\_LSTM\_RF\_CMX1\_AL\_CP1.csv | 0.99949131 | 0.96261733 | 0.02531709 | 0.21631345 |
| NonPeak\_LSTM\_RF\_CMX1\_AL\_CP1.csv | 0.99876519 | 0.75385667 | 0.01127585 | 0.28989043 |
| Peak Aware Model – Z-Score |  |  |  |  |

**CMX1\_AL\_CP2.csv :-**

# of time points – 74822

Step 1 – Global Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| CMX1\_AL\_CP2.csv | 0.9988 | -1.0846 | 0.0300 | 1.1650 |

Step 2 – Peak Detection

Autoencoder # of Peaks – 14281 [Window\_Size = 30, Step = 5, Latent\_Space = 16]

Z-Score # of Peaks – 26203 [Lag = 5000, Influence=0.5, Threshold =0.5xMean]

Step 3 – Statistical Analysis

For Autoencoder

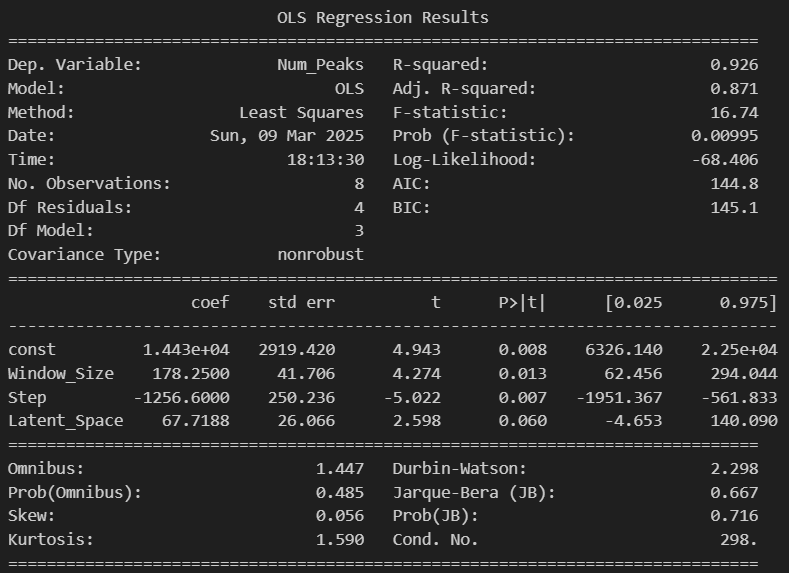
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Window\_Size) | 57191512.5 | 1 | 5.486429 | 0.25687795 |
| C(Step) | 78952178 | 1 | 7.57394761 | 0.22187993 |
| C(Latent\_Space) | 21131500.5 | 1 | 2.02716229 | 0.38980467 |
| C(Window\_Size):C(Step) | 758912 | 1 | 0.07280305 | 0.8322224 |
| C(Window\_Size):C(Latent\_Space) | 1144584.5 | 1 | 0.10980094 | 0.79629725 |
| C(Step):C(Latent\_Space) | 195938 | 1 | 0.01879649 | 0.91326001 |
| Residual | 10424178 | 1 |  |  |

1. Tukey’s Test

Not Applicable since no significant features were found.

1. Multinomial Regression



For Z-Score

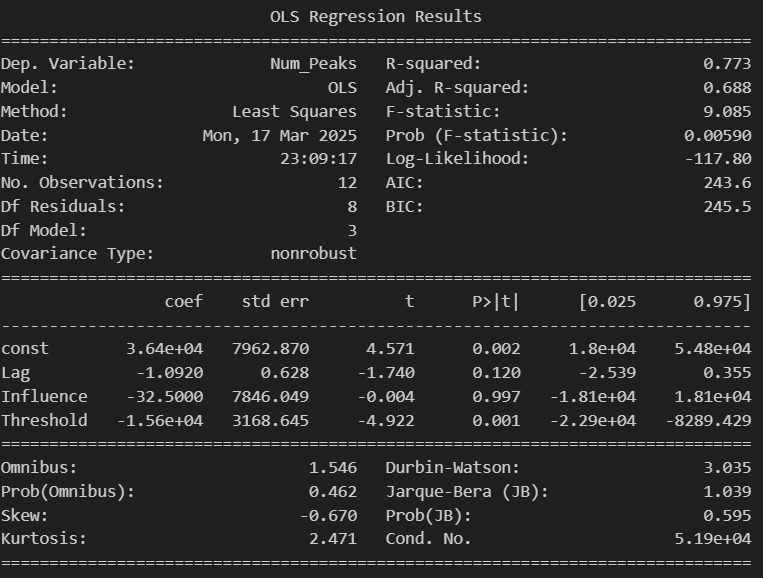
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Lag) | 89434800 | 1 | 4345010.53 | 2.30E-07 |
| C(Influence) | 507 | 1 | 24.6315789 | 0.03828217 |
| C(Threshold) | 902670750 | 2 | 21927224.7 | 4.56E-08 |
| C(Lag):C(Influence) | 21.3333333 | 1 | 1.03643725 | 0.41576261 |
| C(Lag):C(Threshold) | 49602664.5 | 2 | 1204923.02 | 8.30E-07 |
| C(Influence):C(Threshold) | 708.5 | 2 | 17.2105263 | 0.05491329 |
| Residual | 41.1666667 | 2 |  |  |

1. Tukey’s Test

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group1 | group2 | meandiff | p-adj | lower | upper | reject | Factor |
| 5000 | 10000 | -5460 | 0.3553 | -18013.4398 | 7093.4398 | False | Lag |
| 0.5 | 0.9 | -13 | 0.9983 | -13142.7008 | 13116.7008 | False | Influence |
| 0.39706782 | 0.79413565 | -15324.5 | 0.001 | -23084.2539 | -7564.7461 | True | Threshold |
| 0.39706782 | 1.5882713 | -20404.75 | 0.0001 | -28164.5039 | -12644.9961 | True | Threshold |
| 0.79413565 | 1.5882713 | -5080.25 | 0.2151 | -12840.0039 | 2679.5039 | False | Threshold |
| 5000 | 10000 | -5460 | 0.3553 | -18013.4398 | 7093.4398 | False | Lag |

1. Multinomial Regression



Step 4 – Classification (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | train\_accuracy | test\_accuracy | train\_f1 | test\_f1 |
| Data2\_Classes.csv RF AE | 0.99993317 | 0.78116542 | 0.99993316 | 0.68583194 |
| Data2\_Classes.csv RF Z-Score | 1 | 0.6332821 | 1 | 0.50715895 |

Step 4 – Regression (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| Peak\_LSTM\_RF\_CMX1\_AL\_CP2.csv | 0.99693455 | 0.03469395 | 0.09388279 | 0.99579176 |
| NonPeak\_LSTM\_RF\_CMX1\_AL\_CP2.csv | 0.99790293 | -1.3818587 | 0.03213168 | 1.07779892 |
|  |  |  |  |  |
| Peak\_LSTM\_RF\_CMX1\_AL\_CP2.csv | 0.99951176 | 0.71448464 | 0.03621512 | 1.19412069 |
| NonPeak\_LSTM\_RF\_CMX1\_AL\_CP2.csv | 0.99982485 | -0.2567748 | 0.00623357 | 0.52201921 |
|  |  |  |  |  |

**CMX1\_S\_CP1.csv :-**

# of time points – 244772

Step 1 – Global Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| CMX1\_S\_CP1.csv | 1 | 0.8420 | 0.0119 | 1.0510 |

Step 2 – Peak Detection

Autoencoder # of Peaks – 40368 [Window\_Size = 30, Step = 5, Latent\_Space = 16]

Z-Score # of Peaks – 66581 [Lag = 5000, Influence=0.5, Threshold =0.5xMean]

Step 3 – Statistical Analysis

For Autoencoder

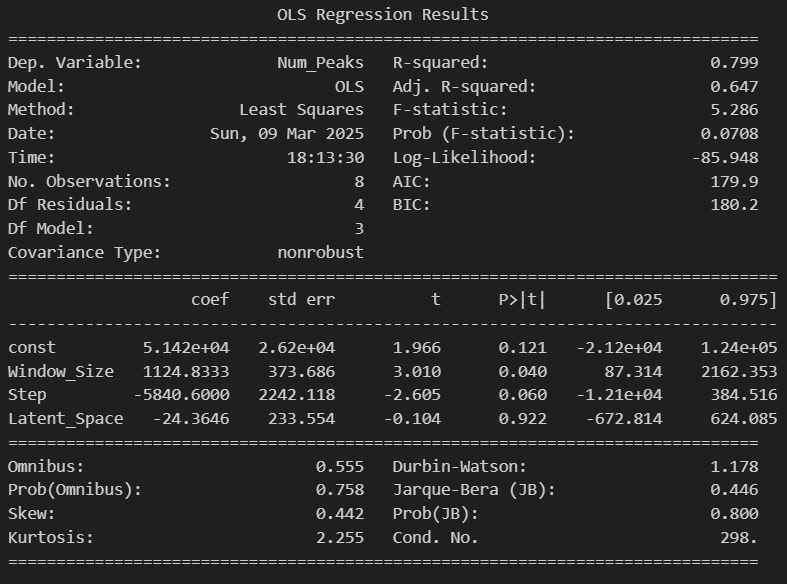
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Window\_Size) | 2277450050 | 1 | 9.05923285 | 0.20420717 |
| C(Step) | 1705630418 | 1 | 6.78465072 | 0.23336219 |
| C(Latent\_Space) | 2735460.5 | 1 | 0.01088111 | 0.93383188 |
| C(Window\_Size):C(Step) | 534841218 | 1 | 2.12748953 | 0.38260285 |
| C(Window\_Size):C(Latent\_Space) | 150771613 | 1 | 0.59973879 | 0.58049772 |
| C(Step):C(Latent\_Space) | 68409904.5 | 1 | 0.27212068 | 0.69390131 |
| Residual | 251395465 | 1 |  |  |

1. Tukey’s Test

Not Applicable since no significant features were found.

1. Multinomial Regression



For Z-Score

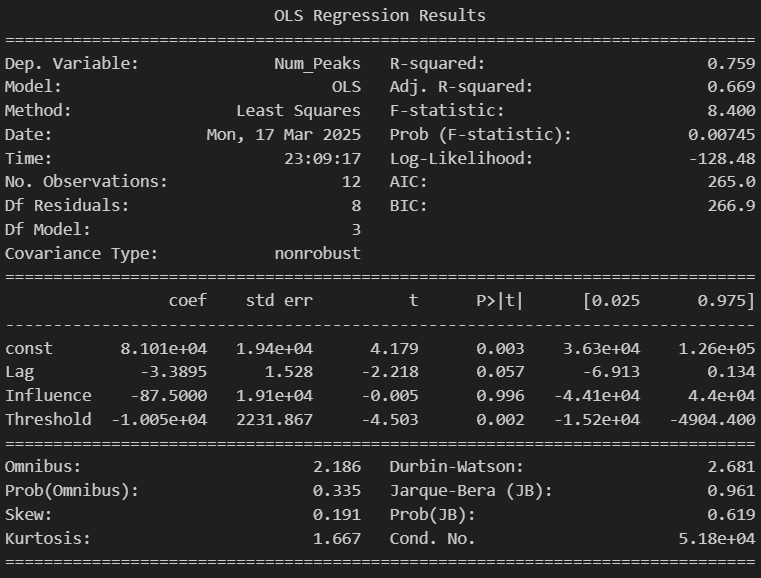
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Lag) | 861636321 | 1 | 759764.557 | 1.32E-06 |
| C(Influence) | 3675 | 1 | 3.24050261 | 0.21364307 |
| C(Threshold) | 4522764018 | 2 | 1994017.5 | 5.01E-07 |
| C(Lag):C(Influence) | 2133.33333 | 1 | 1.88110809 | 0.30380804 |
| C(Lag):C(Threshold) | 430239286 | 2 | 189685.922 | 5.27E-06 |
| C(Influence):C(Threshold) | 3106.5 | 2 | 1.36960835 | 0.42201067 |
| Residual | 2268.16667 | 2 |  |  |

1. Tukey’s Test

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group1 | group2 | meandiff | p-adj | lower | upper | reject | Factor |
| 5000 | 10000 | -16947.3333 | 0.2166 | -45577.0087 | 11682.342 | False | Lag |
| 1.37241462 | 2.74482923 | -34612.75 | 0.0069 | -58266.0785 | -10959.4215 | True | Threshold |
| 1.37241462 | 5.48965847 | -45546.5 | 0.0012 | -69199.8285 | -21893.1715 | True | Threshold |
| 2.74482923 | 5.48965847 | -10933.75 | 0.4347 | -34587.0785 | 12719.5785 | False | Threshold |
| 5000 | 10000 | -16947.3333 | 0.2166 | -45577.0087 | 11682.342 | False | Lag |

1. Multinomial Regression



Step 4 – Classification (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | train\_accuracy | test\_accuracy | train\_f1 | test\_f1 |
| Data3\_Classes.csv RF AE | 1 | 0.85399197 | 1 | 0.81659781 |
| Data3\_Classes.csv RF Z-Score | 1 | 0.71519472 | 1 | 0.63909272 |

Step 5 – Regression (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| Peak\_LSTM\_RF\_CMX1\_S\_CP1.csv | 0.99986528 | 0.85312377 | 0.03754915 | 0.9604815 |
| NonPeak\_LSTM\_RF\_CMX1\_S\_CP1.csv | 0.99992541 | 0.96131821 | 0.03285655 | 0.49887586 |
|  |  |  |  |  |
| Peak\_LSTM\_RF\_CMX1\_S\_CP1.csv | 0.99978243 | 0.94026918 | 0.05432456 | 0.16872649 |
| NonPeak\_LSTM\_RF\_CMX1\_S\_CP1.csv | 0.9991934 | 0.96400862 | 0.1138229 | 0.49328116 |
|  |  |  |  |  |

**CMX1\_S\_CP2.csv :-**

# of time points – 45468

Step 1 – Global Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| CMX1\_S\_CP2.csv | 0.9999 | -0.3814 | 0.0142 | 1.5052 |

Step 2 – Peak Detection

Autoencoder # of Peaks – 7020 [Window\_Size = 30, Step = 5, Latent\_Space = 16]

Z-Score # of Peaks – 17008 [Lag = 5000, Influence=0.5, Threshold =0.5xMean]

Step 3 – Statistical Analysis

For Autoencoder

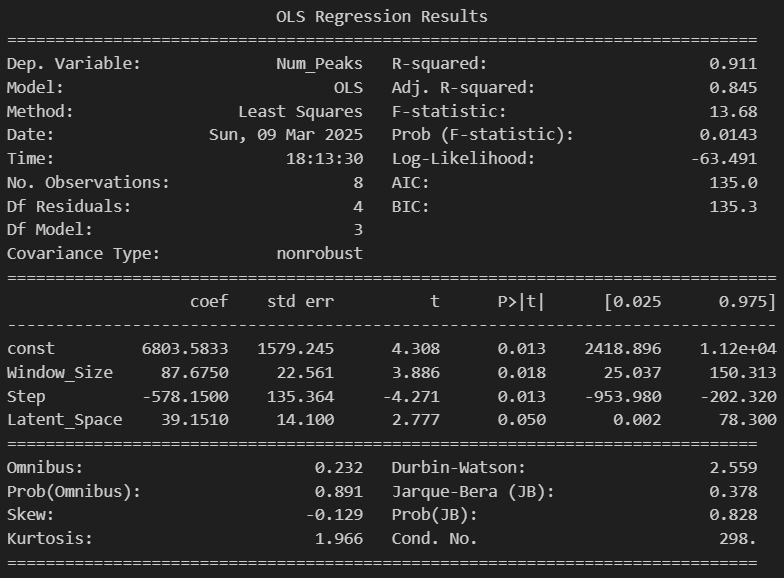
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Window\_Size) | 13836430.1 | 1 | 5.48328986 | 0.25694375 |
| C(Step) | 16712871.1 | 1 | 6.62320525 | 0.2359387 |
| C(Latent\_Space) | 7063161.12 | 1 | 2.79908614 | 0.34296957 |
| C(Window\_Size):C(Step) | 82621.125 | 1 | 0.03274223 | 0.88603795 |
| C(Window\_Size):C(Latent\_Space) | 473851.125 | 1 | 0.18778421 | 0.73967721 |
| C(Step):C(Latent\_Space) | 584821.125 | 1 | 0.23176092 | 0.71436793 |
| Residual | 2523381.13 | 1 |  |  |

1. Tukey’s Test

Not Applicable since no significant features were found.

1. Multinomial Regression



For Z-Score

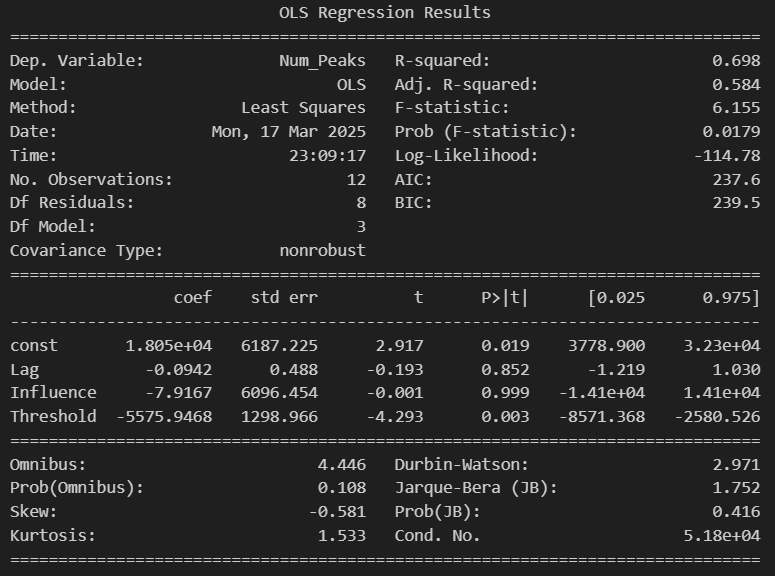
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Lag) | 665994.083 | 1 | 61952.938 | 1.61E-05 |
| C(Influence) | 30.0833333 | 1 | 2.79844961 | 0.23632552 |
| C(Threshold) | 470178238 | 2 | 21868755.3 | 4.57E-08 |
| C(Lag):C(Influence) | 18.75 | 1 | 1.74418605 | 0.31747637 |
| C(Lag):C(Threshold) | 1270676.17 | 2 | 59101.2171 | 1.69E-05 |
| C(Influence):C(Threshold) | 4.16666667 | 2 | 0.19379845 | 0.83766234 |
| Residual | 21.5 | 2 |  |  |

1. Tukey’s Test

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group1 | group2 | meandiff | p-adj | lower | upper | reject | Factor |
| 5000 | 10000 | -471.1667 | 0.9077 | -9303.9785 | 8361.6452 | False | Lag |
| 0.7526044 | 1.50520881 | -12143 | 0 | -13058.8329 | -11227.1671 | True | Threshold |
| 0.7526044 | 3.01041762 | -14178.75 | 0 | -15094.5829 | -13262.9171 | True | Threshold |
| 1.50520881 | 3.01041762 | -2035.75 | 0.0004 | -2951.5829 | -1119.9171 | True | Threshold |
| 5000 | 10000 | -471.1667 | 0.9077 | -9303.9785 | 8361.6452 | False | Lag |

1. Multinomial Regression



Step 4 – Classification (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | train\_accuracy | test\_accuracy | train\_f1 | test\_f1 |
| Data4\_Classes.csv RF AE | 0.99978005 | 0.82617182 | 0.99977987 | 0.75135447 |
| Data4\_Classes.csv RF Z-Score | 1 | 0.62347766 | 1 | 0.53101417 |

Step 4 – Regression (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| Peak\_LSTM\_RF\_CMX1\_S\_CP2.csv | 0.99697988 | -2.3836755 | 0.09106114 | 0.88439624 |
| NonPeak\_LSTM\_RF\_CMX1\_S\_CP2.csv | 0.99822859 | -0.9492774 | 0.04454153 | 1.37178916 |
|  |  |  |  |  |
| Peak\_LSTM\_RF\_CMX1\_S\_CP2.csv | 0.99659107 | -6.3718266 | 0.01035213 | 0.60113854 |
| NonPeak\_LSTM\_RF\_CMX1\_S\_CP2.csv | 0.99957385 | 0.03774121 | 0.02574691 | 0.97994995 |
|  |  |  |  |  |

**DMC2\_AL\_CP1.csv :-**

# of time points – 510050

Step 1 – Global Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| DMC2\_AL\_CP1.csv | 0.9990 | -4.4139 | 0.0344 | 4.1321 |

Step 2 – Peak Detection

Autoencoder # of Peaks – 80544 [Window\_Size = 30, Step = 5, Latent\_Space = 16]

Z-Score # of Peaks – 286205 [Lag = 5000, Influence=0.5, Threshold =0.5xMean]

Step 3 – Statistical Analysis

For Autoencoder

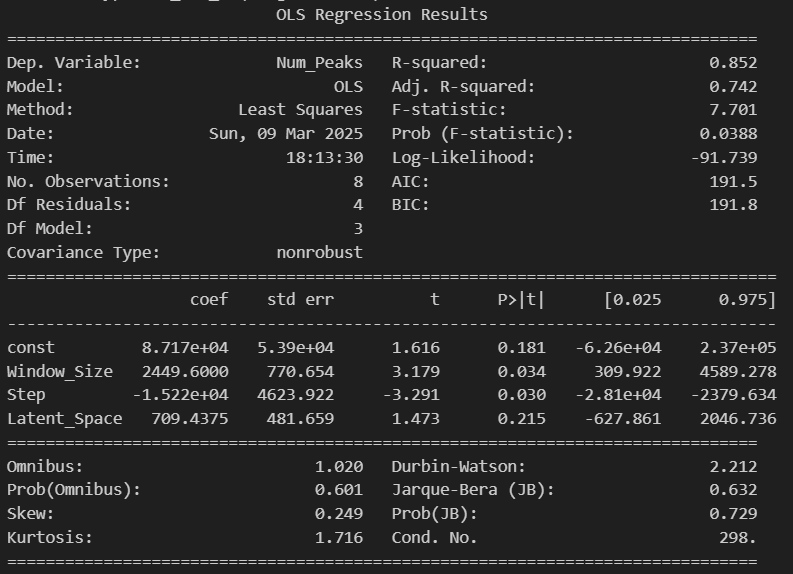
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Window\_Size) | 1.0801E+10 | 1 | 168.013174 | 0.04901726 |
| C(Step) | 1.1579E+10 | 1 | 180.114437 | 0.04734826 |
| C(Latent\_Space) | 2319213618 | 1 | 36.0762375 | 0.10502777 |
| C(Window\_Size):C(Step) | 3319006338 | 1 | 51.6283882 | 0.08803495 |
| C(Window\_Size):C(Latent\_Space) | 844974940 | 1 | 13.1439021 | 0.171337 |
| C(Step):C(Latent\_Space) | 47863328 | 1 | 0.74453202 | 0.54678129 |
| Residual | 64286460.5 | 1 |  |  |

1. Tukey’s Test

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group1 | group2 | meandiff | p-adj | lower | upper | reject | Factor |
| 30 | 60 | 73488 | 0.1079 | -21738.1271 | 168714.127 | False | Window\_Size |
| 5 | 10 | -76088.5 | 0.0926 | -169254.266 | 17077.2657 | False | Step |

1. Multinomial Regression



For Z-Score

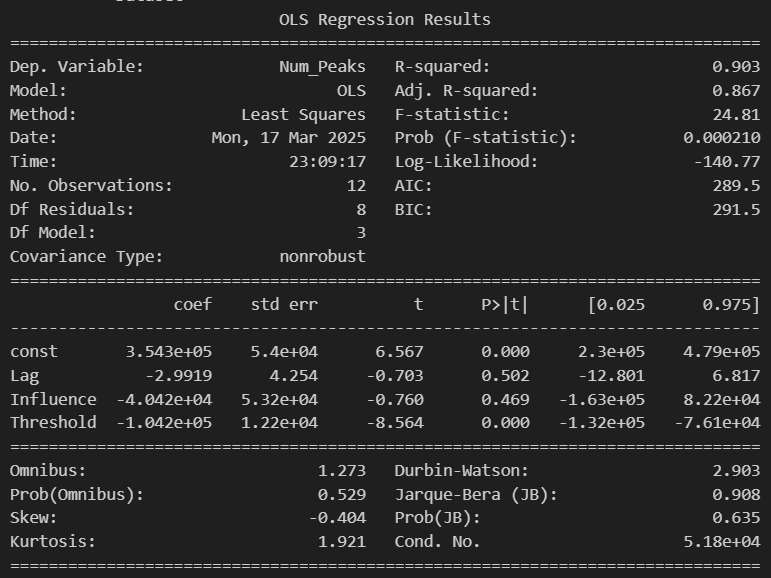
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Lag) | 671359921 | 1 | 92.3921787 | 0.01065082 |
| C(Influence) | 784034834 | 1 | 107.898438 | 0.00914109 |
| C(Threshold) | 1.0869E+11 | 2 | 7478.73319 | 0.00013369 |
| C(Lag):C(Influence) | 23826190.1 | 1 | 3.27894702 | 0.21187782 |
| C(Lag):C(Threshold) | 1110891604 | 2 | 76.4401421 | 0.0129132 |
| C(Influence):C(Threshold) | 548838978 | 2 | 37.7654574 | 0.02579616 |
| Residual | 14532830.2 | 2 |  |  |

1. Tukey’s Test

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group1 | group2 | meandiff | p-adj | lower | upper | reject | Factor |
| 5000 | 10000 | -14959.5 | 0.8108 | -150595.205 | 120676.205 | False | Lag |
| 0.5 | 0.9 | -16166.1667 | 0.7959 | -151733.118 | 119400.784 | False | Influence |
| 0.70076967 | 1.40153934 | -136959.25 | 0 | -173914.435 | -100004.066 | True | Threshold |
| 0.70076967 | 2.80307869 | -231848.25 | 0 | -268803.435 | -194893.066 | True | Threshold |
| 1.40153934 | 2.80307869 | -94889 | 0.0001 | -131844.185 | -57933.8155 | True | Threshold |
| 5000 | 10000 | -14959.5 | 0.8108 | -150595.205 | 120676.205 | False | Lag |
| 0.5 | 0.9 | -16166.1667 | 0.7959 | -151733.118 | 119400.784 | False | Influence |

1. Multinomial Regression



Step 4 – Classification (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | train\_accuracy | test\_accuracy | train\_f1 | test\_f1 |
| Data5\_Classes.csv RF AE | 1 | 0.80511715 | 1 | 0.71822537 |
| Data5\_Classes.csv RF Z-Score | 1 | 0.58387413 | 1 | 0.58162579 |

Step 5 – Regression (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| Peak\_LSTM\_RF\_DMC2\_AL\_CP1.csv | 0.99408898 | 0.20629867 | 0.26980544 | 2.57185229 |
| NonPeak\_LSTM\_RF\_DMC2\_AL\_CP1.csv | 0.99953864 | 0.49254102 | 0.02187426 | 1.24607089 |
|  |  |  |  |  |
| Peak\_LSTM\_RF\_DMC2\_AL\_CP1.csv | 0.99881267 | 0.61748415 | 0.03382183 | 1.14457538 |
| NonPeak\_LSTM\_RF\_DMC2\_AL\_CP1.csv | 0.99948483 | 0.81966885 | 0.02496669 | 0.65420158 |
|  |  |  |  |  |

**DMC2\_AL\_CP2.csv :-**

# of time points – 64337

Step 1 – Global Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| DMC2\_AL\_CP2.csv | 0.9955 | -1.2232 | 0.1405 | 4.5284 |

Step 2 – Peak Detection

Autoencoder # of Peaks – 11523 [Window\_Size = 30, Step = 5, Latent\_Space = 16]

Z-Score # of Peaks – 8601 [Lag = 5000, Influence=0.5, Threshold =0.5xMean]

Step 3 – Statistical Analysis

For Autoencoder

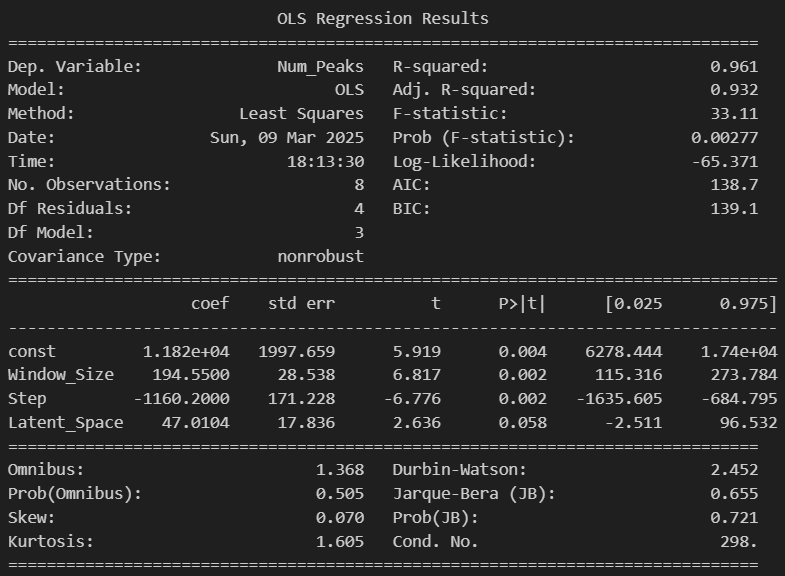
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Window\_Size) | 68129464.5 | 1 | 14.0336316 | 0.16606756 |
| C(Step) | 67303202 | 1 | 13.8634342 | 0.16703796 |
| C(Latent\_Space) | 10183584.5 | 1 | 2.09766325 | 0.38470146 |
| C(Window\_Size):C(Step) | 539760.5 | 1 | 0.11118244 | 0.79510595 |
| C(Window\_Size):C(Latent\_Space) | 165888 | 1 | 0.0341704 | 0.88363289 |
| C(Step):C(Latent\_Space) | 303420.5 | 1 | 0.0625 | 0.84404174 |
| Residual | 4854728 | 1 |  |  |

1. Tukey’s Test

Not Applicable since no significant features were found.

1. Multinomial Regression



For Z-Score

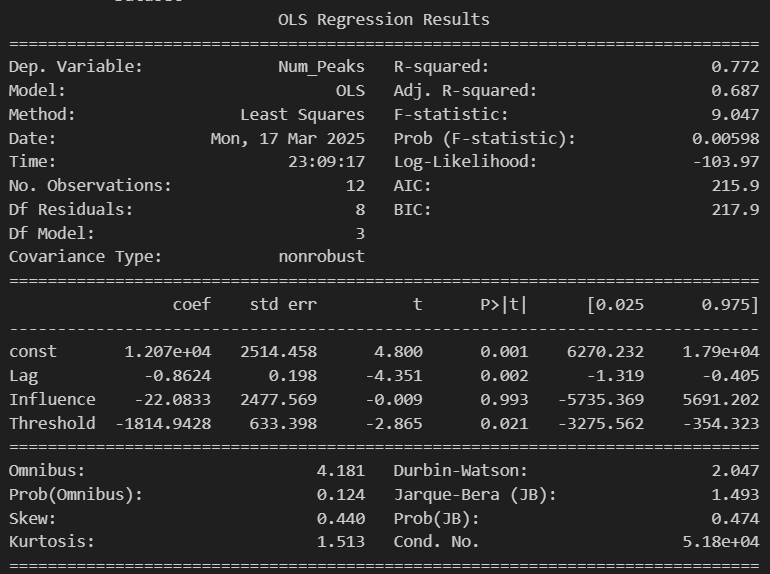
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Lag) | 55775720.1 | 1 | 3084371.62 | 3.24E-07 |
| C(Influence) | 234.083333 | 1 | 12.9447005 | 0.06931568 |
| C(Threshold) | 25373361.2 | 2 | 701567.59 | 1.43E-06 |
| C(Lag):C(Influence) | 154.083333 | 1 | 8.52073733 | 0.10005597 |
| C(Lag):C(Threshold) | 22389291.2 | 2 | 619058.742 | 1.62E-06 |
| C(Influence):C(Threshold) | 28.1666667 | 2 | 0.77880184 | 0.56217617 |
| Residual | 36.1666667 | 2 |  |  |

1. Tukey’s Test

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group1 | group2 | meandiff | p-adj | lower | upper | reject | Factor |
| 5000 | 10000 | -4311.8333 | 0.0066 | -7123.2674 | -1500.3993 | True | Lag |
| 0.62724331 | 1.25448662 | -1864.75 | 0.6568 | -7682.9345 | 3953.4345 | False | Threshold |
| 0.62724331 | 2.50897325 | -3560.5 | 0.2542 | -9378.6845 | 2257.6845 | False | Threshold |
| 1.25448662 | 2.50897325 | -1695.75 | 0.7043 | -7513.9345 | 4122.4345 | False | Threshold |
| 5000 | 10000 | -4311.8333 | 0.0066 | -7123.2674 | -1500.3993 | True | Lag |

1. Multinomial Regression



Step 4 – Classification (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | train\_accuracy | test\_accuracy | train\_f1 | test\_f1 |
| Data6\_Classes.csv RF AE | 0.99945597 | 0.82904605 | 0.99945621 | 0.77205699 |
| Data6\_Classes.csv RF Z-Score | 1 | 1 | 1 | 1 |

Step 4 – Regression (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| Peak\_LSTM\_RF\_DMC2\_AL\_CP2.csv | 0.99792578 | 0.7629732 | 0.26885356 | 6.80121373 |
| NonPeak\_LSTM\_RF\_DMC2\_AL\_CP2.csv | 0.98917753 | -0.1784011 | 0.03426125 | 0.41358364 |
|  |  |  |  |  |
| Peak\_LSTM\_RF\_DMC2\_AL\_CP2.csv | 0.99887868 | 0.57604889 | 0.03286817 | 1.2049737 |
| NonPeak\_LSTM\_RF\_DMC2\_AL\_CP2.csv | 0.99785446 | 0.80374707 | 0.05095135 | 0.68247115 |
|  |  |  |  |  |

**DMC2\_S\_CP1.csv :-**

# of time points – 263913

Step 1 – Global Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| DMC2\_S\_CP1.csv | 0.9997 | 0.1944 | 0.0590 | 2.8296 |

Step 2 – Peak Detection

Autoencoder # of Peaks – 55899 [Window\_Size = 30, Step = 5, Latent\_Space = 16]

Z-Score # of Peaks – 70532 [Lag = 5000, Influence=0.5, Threshold =0.5xMean]

Step 3 – Statistical Analysis

For Autoencoder

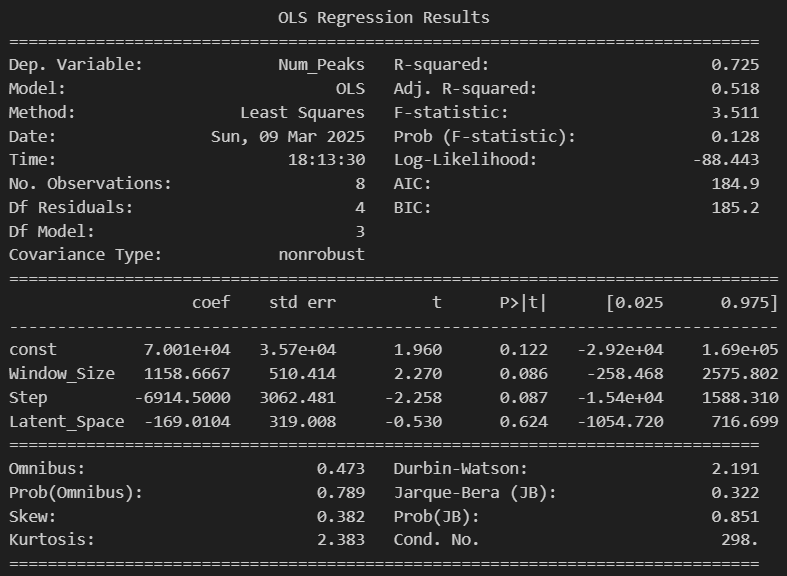
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Window\_Size) | 2416515200 | 1 | 3.74665411 | 0.30357952 |
| C(Step) | 2390515513 | 1 | 3.70634324 | 0.30498585 |
| C(Latent\_Space) | 131625313 | 1 | 0.20407673 | 0.72987774 |
| C(Window\_Size):C(Step) | 81447084.5 | 1 | 0.12627856 | 0.78263308 |
| C(Window\_Size):C(Latent\_Space) | 535921061 | 1 | 0.83091174 | 0.52943849 |
| C(Step):C(Latent\_Space) | 613410338 | 1 | 0.95105397 | 0.50798627 |
| Residual | 644979528 | 1 |  |  |

1. Tukey’s Test

Not Applicable since no significant features were found.

1. Multinomial Regression



For Z-Score

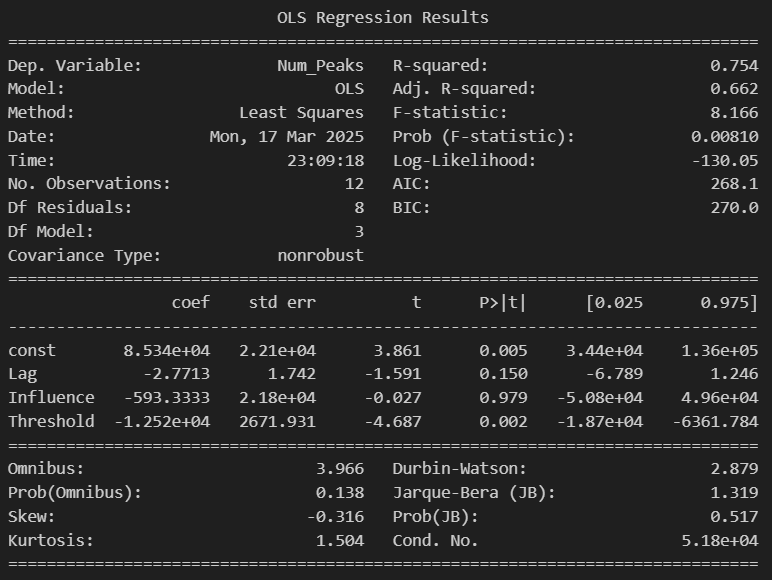
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Lag) | 575993920 | 1 | 7561.59675 | 0.00013222 |
| C(Influence) | 168981.333 | 1 | 2.21837185 | 0.2748218 |
| C(Threshold) | 6632650840 | 2 | 43536.424 | 2.30E-05 |
| C(Lag):C(Influence) | 93280.3333 | 1 | 1.22457589 | 0.38375054 |
| C(Lag):C(Threshold) | 188448408 | 2 | 1236.96694 | 0.00080778 |
| C(Influence):C(Threshold) | 236561.167 | 2 | 1.55277694 | 0.39173027 |
| Residual | 152347.167 | 2 |  |  |

1. Tukey’s Test

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group1 | group2 | meandiff | p-adj | lower | upper | reject | Factor |
| 5000 | 10000 | -13856.3333 | 0.3798 | -47455.5798 | 19742.9131 | False | Lag |
| 1.30694433 | 2.61388867 | -43360.25 | 0.0002 | -61563.012 | -25157.488 | True | Threshold |
| 1.30694433 | 5.22777733 | -54500.25 | 0 | -72703.012 | -36297.488 | True | Threshold |
| 2.61388867 | 5.22777733 | -11140 | 0.2542 | -29342.762 | 7062.762 | False | Threshold |
| 5000 | 10000 | -13856.3333 | 0.3798 | -47455.5798 | 19742.9131 | False | Lag |

1. Multinomial Regression



Step 4 – Classification (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | train\_accuracy | test\_accuracy | train\_f1 | test\_f1 |
| Data7\_Classes.csv RF AE | 0.99986738 | 0.78853413 | 0.99986738 | 0.69995713 |
| Data7\_Classes.csv RF Z-Score | 1 | 0.68049694 | 1 | 0.66407674 |

Step 5 – Regression (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| Peak\_LSTM\_RF\_DMC2\_S\_CP1.csv | 0.99874423 | 0.95380422 | 0.10220201 | 0.83985083 |
| NonPeak\_LSTM\_RF\_DMC2\_S\_CP1.csv | 0.9995152 | 0.71670277 | 0.07183509 | 1.60994073 |
|  |  |  |  |  |
| Peak\_LSTM\_RF\_DMC2\_S\_CP1.csv | 0.99886482 | 0.98591179 | 0.09860361 | 0.30676019 |
| NonPeak\_LSTM\_RF\_DMC2\_S\_CP1.csv | 0.99962714 | 0.68167578 | 0.06715547 | 1.73869231 |
|  |  |  |  |  |

**DMC2\_S\_CP2.csv :-**

# of time points – 42016

Step 1 – Global Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| DMC2\_S\_CP2.csv | 0.9901 | -0.1073 | 0.2122 | 3.0129 |

Step 2 – Peak Detection

Autoencoder # of Peaks – 7088 [Window\_Size = 30, Step = 5, Latent\_Space = 16]

Z-Score # of Peaks – 2426 [Lag = 5000, Influence=0.5, Threshold =0.5xMean]

Step 3 – Statistical Analysis

For Autoencoder

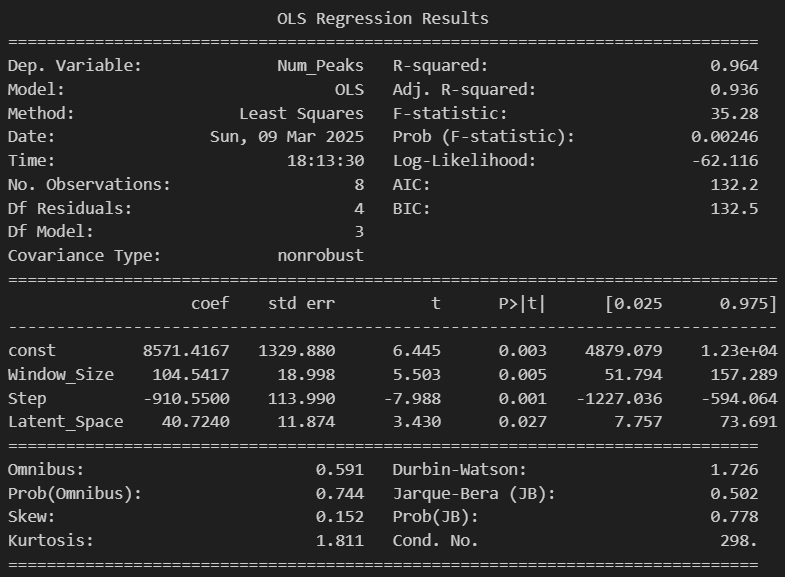
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Window\_Size) | 19672128.1 | 1 | 38.9150139 | 0.10119109 |
| C(Step) | 41455065.1 | 1 | 82.0055881 | 0.07001682 |
| C(Latent\_Space) | 7642095.12 | 1 | 15.1174411 | 0.16026136 |
| C(Window\_Size):C(Step) | 1935528.13 | 1 | 3.82882337 | 0.30077239 |
| C(Window\_Size):C(Latent\_Space) | 98346.125 | 1 | 0.19454636 | 0.73554408 |
| C(Step):C(Latent\_Space) | 59340.125 | 1 | 0.11738546 | 0.78986377 |
| Residual | 505515.125 | 1 |  |  |

1. Tukey’s Test

Not Applicable since no significant features were found.

1. Multinomial Regression



For Z-Score

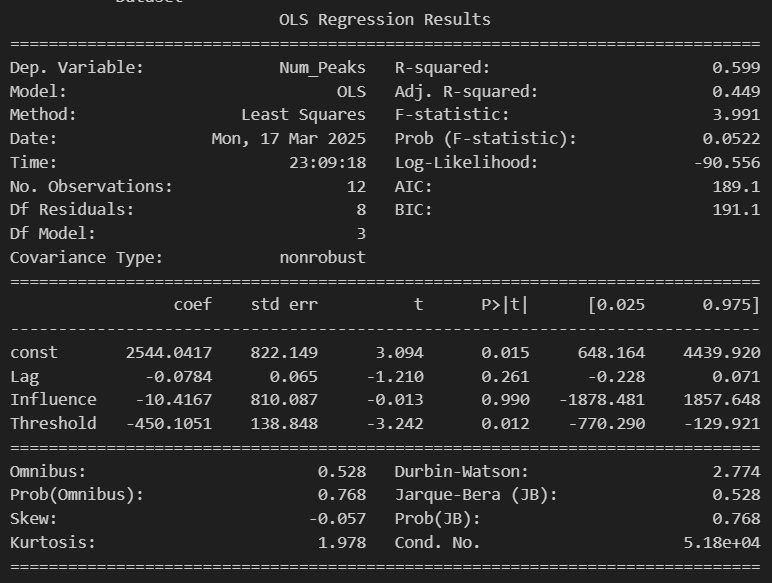
1. ANOVA test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | sum\_sq | df | F | PR(>F) |
| C(Lag) | 461384.083 | 1 | 291400.474 | 3.43E-06 |
| C(Influence) | 52.0833333 | 1 | 32.8947368 | 0.02908045 |
| C(Threshold) | 5173873.17 | 2 | 1633854.68 | 6.12E-07 |
| C(Lag):C(Influence) | 4.08333333 | 1 | 2.57894737 | 0.24952123 |
| C(Lag):C(Threshold) | 656187.167 | 2 | 207217 | 4.83E-06 |
| C(Influence):C(Threshold) | 81.1666667 | 2 | 25.6315789 | 0.03754941 |
| Residual | 3.16666667 | 2 |  |  |

1. Tukey’s Test

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| group1 | group2 | meandiff | p-adj | lower | upper | reject | Factor |
| 5000 | 10000 | -392.1667 | 0.3946 | -1374.4197 | 590.0864 | False | Lag |
| 0.5 | 0.9 | -4.1667 | 0.9929 | -1024.5418 | 1016.2084 | False | Influence |
| 0.9355751 | 1.8711502 | -1333.25 | 0.0012 | -2028.9867 | -637.5133 | True | Threshold |
| 0.9355751 | 3.7423004 | -1445.75 | 0.0007 | -2141.4867 | -750.0133 | True | Threshold |
| 1.8711502 | 3.7423004 | -112.5 | 0.895 | -808.2367 | 583.2367 | False | Threshold |
| 5000 | 10000 | -392.1667 | 0.3946 | -1374.4197 | 590.0864 | False | Lag |
| 0.5 | 0.9 | -4.1667 | 0.9929 | -1024.5418 | 1016.2084 | False | Influence |

1. Multinomial Regression



Step 4 – Classification (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | train\_accuracy | test\_accuracy | train\_f1 | test\_f1 |
| Data8\_Classes.csv RF AE | 0.99738189 | 0.83063101 | 0.99739332 | 0.76494041 |
| Data8\_Classes.csv RF Z-Score | 1 | 0.92940231 | 1 | 0.89539507 |

Step 5 – Regression (Split 20-80)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Train\_R2 | Test\_R2 | Train\_RMSE | Test\_RMSE |
| Peak\_LSTM\_RF\_DMC2\_S\_CP2.csv | 0.99750195 | 0.67561466 | 0.28406452 | 5.75196338 |
| NonPeak\_LSTM\_RF\_DMC2\_S\_CP2.csv | 0.99389993 | 0.19326693 | 0.05882409 | 1.92142977 |
|  |  |  |  |  |
| Peak\_LSTM\_RF\_DMC2\_S\_CP2.csv | 0.99875601 | 0.97717748 | 0.10322091 | 0.39043875 |
| NonPeak\_LSTM\_RF\_DMC2\_S\_CP2.csv | 0.99961915 | 0.66988879 | 0.0678717 | 1.77059007 |
|  |  |  |  |  |