

ARIMA Model for predicting Daily New COVID19 Cases – Model Accuracy Report

By

Surya Vaddhiparthy

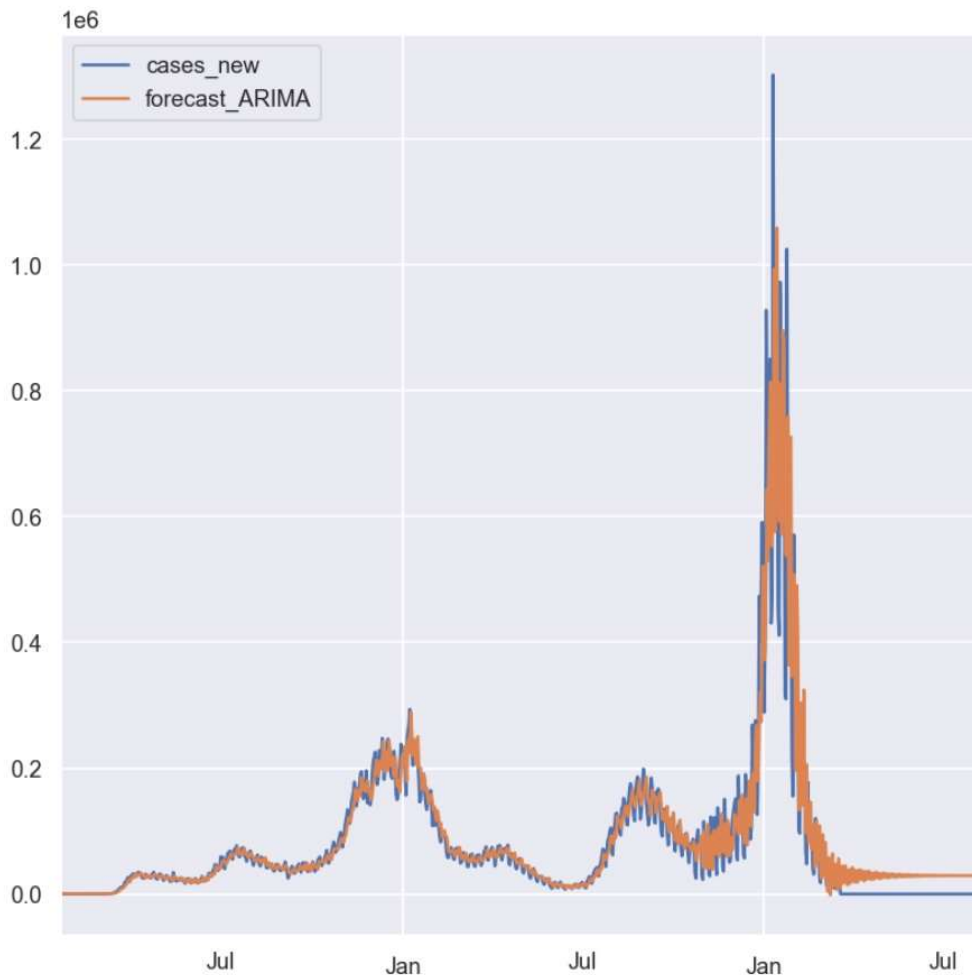
Under the guidance of

Professor Dr. Hyun Bin Kang

1. Dataset modifications to improve accuracy:

The dataset has now been changed from weekly resolution to daily data yielding 787 rows one for each day from the start of the pandemic.

2. The output of ARIMA (2,1,3) Model:



The orange line shows the predicted values of the ARIMA Model, this dataset contains daily values until 3/18/2022 and the rest are zeroes for the blue line which indicates the data for those dates (after 3/18/22) is not fed into the model. It may not seem like a significant extrapolation, but further analysis yields interesting results of the tuned ARIMA (2,1,3) model. The tuning part has been added to the further reading section at the end.

These values towards the end of the plot are the new values after the end of peak COVID Cases that took place during the end of January and model has incorporated those values into its computation process and the data was fed until the end of the peak and few days extra days when of COVID new case count fell.

3. ARIMA Model Summary:

| SARIMAX Results | | | | | | |
|-------------------------|------------------|-------------------|-------------------|-------|-----------|----------|
| Dep. Variable: | cases_new | | No. Observations: | | 787 | |
| Model: | ARIMA(2, 1, 3) | | Log Likelihood | | -9638.465 | |
| Date: | Mon, 28 Mar 2022 | | AIC | | 19288.930 | |
| Time: | 22:47:01 | | BIC | | 19316.931 | |
| Sample: | 01-22-2020 | | HQIC | | 19299.695 | |
| | - 03-18-2022 | | | | | |
| Covariance Type: | opg | | | | | |
| | coef | std err | z | P> z | [0.025 | 0.975] |
| ar.L1 | -0.5038 | 0.007 | -74.777 | 0.000 | -0.517 | -0.491 |
| ar.L2 | -0.9362 | 0.007 | -132.528 | 0.000 | -0.950 | -0.922 |
| ma.L1 | 0.0827 | 0.014 | 5.991 | 0.000 | 0.056 | 0.110 |
| ma.L2 | 0.4486 | 0.015 | 29.936 | 0.000 | 0.419 | 0.478 |
| ma.L3 | -0.5493 | 0.016 | -35.143 | 0.000 | -0.580 | -0.519 |
| sigma2 | 2.929e+09 | 1.31e-12 | 2.24e+21 | 0.000 | 2.93e+09 | 2.93e+09 |
| Ljung-Box (L1) (Q): | 0.45 | Jarque-Bera (JB): | 49755.50 | | | |
| Prob(Q): | 0.50 | Prob(JB): | 0.00 | | | |
| Heteroskedasticity (H): | 309.72 | Skew: | 2.47 | | | |
| Prob(H) (two-sided): | 0.00 | Kurtosis: | 41.66 | | | |

4. Predictions and Accuracy:

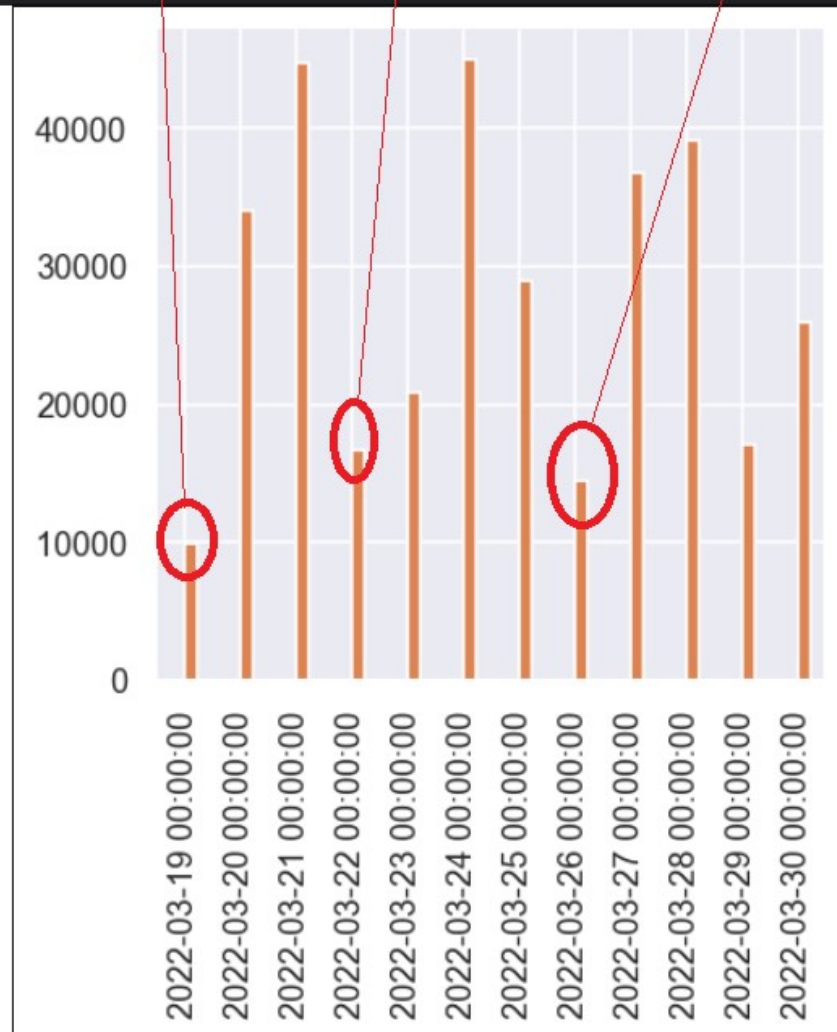
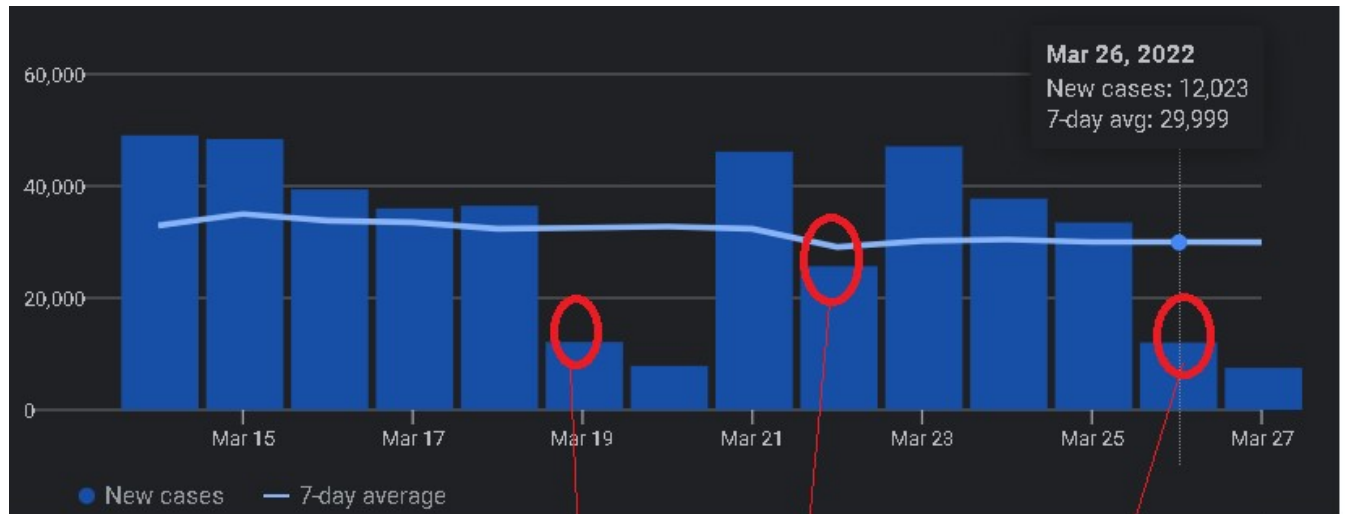
ARIMA (2, 1, 3) has RMSE=94874.740 (for data with an order of 10^6), and below are the values of predictions from unseen data by the model.

| | cases_new | forecast_ARIMA |
|------------|-----------|----------------|
| 2022-03-19 | 0 | 9876.149889 |
| 2022-03-20 | 0 | 34086.353168 |
| 2022-03-21 | 0 | 44776.290649 |
| 2022-03-22 | 0 | 16724.378579 |
| 2022-03-23 | 0 | 20848.637971 |
| 2022-03-24 | 0 | 45033.907564 |
| 2022-03-25 | 0 | 28988.148654 |
| 2022-03-26 | 0 | 14428.981333 |
| 2022-03-27 | 0 | 36786.400690 |
| 2022-03-28 | 0 | 39153.509873 |
| 2022-03-29 | 0 | 17029.258808 |
| 2022-03-30 | 0 | 25959.260503 |

5. Correlation with real-world data:

The model does not know about COVID cases after 18th March, and this is how its predictions fared against the real-world data. The top graph is from google with real-world data and the below is from my predicted values of the ARIMA model.

It is interesting to note the model predicted the dip in cases compared to the neighboring days, on March 19th, March 22nd, and March 26th, and the value was close to the real world 7-day average of 30,000.



Further Reading

ARIMA Model hyperparameter tuning process.

```
ARIMA(0, 0, 0) RMSE=227104.054
ARIMA(0, 0, 1) RMSE=152920.401
ARIMA(0, 0, 2) RMSE=147597.264
ARIMA(0, 0, 3) RMSE=134076.860
ARIMA(0, 0, 4) RMSE=121937.608
ARIMA(0, 1, 0) RMSE=119283.524
ARIMA(0, 1, 1) RMSE=107080.136
ARIMA(0, 1, 2) RMSE=105776.182
ARIMA(0, 1, 3) RMSE=97657.197
ARIMA(0, 1, 4) RMSE=100079.927
ARIMA(0, 2, 0) RMSE=191627.167
ARIMA(0, 2, 1) RMSE=119597.441
ARIMA(0, 2, 2) RMSE=107336.545
ARIMA(0, 2, 3) RMSE=99614.839
ARIMA(0, 2, 4) RMSE=111985.215
ARIMA(1, 0, 0) RMSE=117339.665
ARIMA(1, 0, 1) RMSE=114994.533
ARIMA(1, 0, 2) RMSE=105514.279
ARIMA(1, 0, 3) RMSE=99898.616
ARIMA(1, 0, 4) RMSE=107923.116
ARIMA(1, 1, 0) RMSE=116583.955
ARIMA(1, 1, 1) RMSE=104530.658
ARIMA(1, 1, 2) RMSE=101838.951
ARIMA(1, 1, 3) RMSE=106711.598
ARIMA(1, 1, 4) RMSE=102931.445
ARIMA(1, 2, 0) RMSE=176488.525
ARIMA(1, 2, 1) RMSE=116928.327
ARIMA(1, 2, 2) RMSE=122133.271
ARIMA(1, 2, 3) RMSE=98868.268
ARIMA(1, 2, 4) RMSE=104090.280
ARIMA(2, 0, 0) RMSE=115350.179
ARIMA(2, 0, 1) RMSE=118948.419
ARIMA(2, 0, 2) RMSE=101692.848
ARIMA(2, 0, 3) RMSE=102439.993
ARIMA(2, 0, 4) RMSE=110074.410
ARIMA(2, 1, 0) RMSE=106356.697
ARIMA(2, 1, 1) RMSE=106064.463
ARIMA(2, 1, 2) RMSE=105113.242
ARIMA(2, 1, 3) RMSE=94874.740
ARIMA(2, 1, 4) RMSE=103487.687
ARIMA(2, 2, 0) RMSE=144193.584
ARIMA(2, 2, 1) RMSE=107438.719
ARIMA(2, 2, 2) RMSE=108037.552
ARIMA(2, 2, 3) RMSE=110915.826
ARIMA(2, 2, 4) RMSE=105560.529
ARIMA(3, 0, 0) RMSE=106514.623
ARIMA(3, 0, 1) RMSE=109793.606
ARIMA(3, 0, 2) RMSE=99335.079
```

ARIMA(3, 0, 3) RMSE=105108.144
ARIMA(3, 1, 0) RMSE=106301.227
ARIMA(3, 1, 1) RMSE=105824.997
ARIMA(3, 1, 2) RMSE=108187.549
ARIMA(3, 1, 3) RMSE=96619.568
ARIMA(3, 1, 4) RMSE=102294.006
ARIMA(3, 2, 0) RMSE=128538.577
ARIMA(3, 2, 1) RMSE=107687.332
ARIMA(3, 2, 2) RMSE=109468.586
ARIMA(3, 2, 3) RMSE=97459.992
Best ARIMA(2, 1, 3) RMSE=94874.740