Day-Level Forecasting for Coronavirus Disease (COVID-19)

Wael K. Hanna, Sadat Academy for Management Sciences, Egypt

Nouran M. Radwan, Sadat Academy for Management Sciences, Egypt*

ABSTRACT

Coronavirus (COVID-19) recently spread quickly all over the world. Most infected people with the coronavirus will experience mild to moderate respiratory illness, but elderly people and those with chronic diseases are more likely to suffer from serious disease, often leading to death. According to the Egyptian Ministry of Health, there are 96,336 confirmed infected cases with coronavirus and 5,141 confirmed deaths from the current outbreak. Accurate forecasting of the spread of confirmed and death cases as well as analysis of the number of infected and deaths are crucially required. The present study aims to explore the usage of support vector machine (SVM) in the prediction of coronavirus infected and death cases in Egypt, which helps in the decision-making process. The forecasting model suggest that the number of coronavirus cases grows exponentially in Egypt and more efforts shall be directed to increase the public awareness with this disease. The proposed method is shown to achieve good accuracy and precision results.

KEYWORDS

Coronavirus, Coronavirus Infected and Death Cases, Data Analysis, Data Forecasting

INTRODUCTION

The support vector machine (SVM) algorithm has showed high efficiency in solving classification issues in medical fields. SVM is data-driven and model-free that is discriminative for prediction in particular cases where sample sizes are small. This technique has been used to improve methods for predicting disease in the clinical setting (Battineni, G., et al., 2019). Support Vector Machine is a discriminative classifier that can be defined by a separating hyper-plane. It is the generalization of maximal margin classifier which comes with the definition of hyper-plane (Islam, M., 2017). SVM is characterized by an optimal hyper-plane to classify new examples and datasets (Gholami R & Fakhari N, 2017).

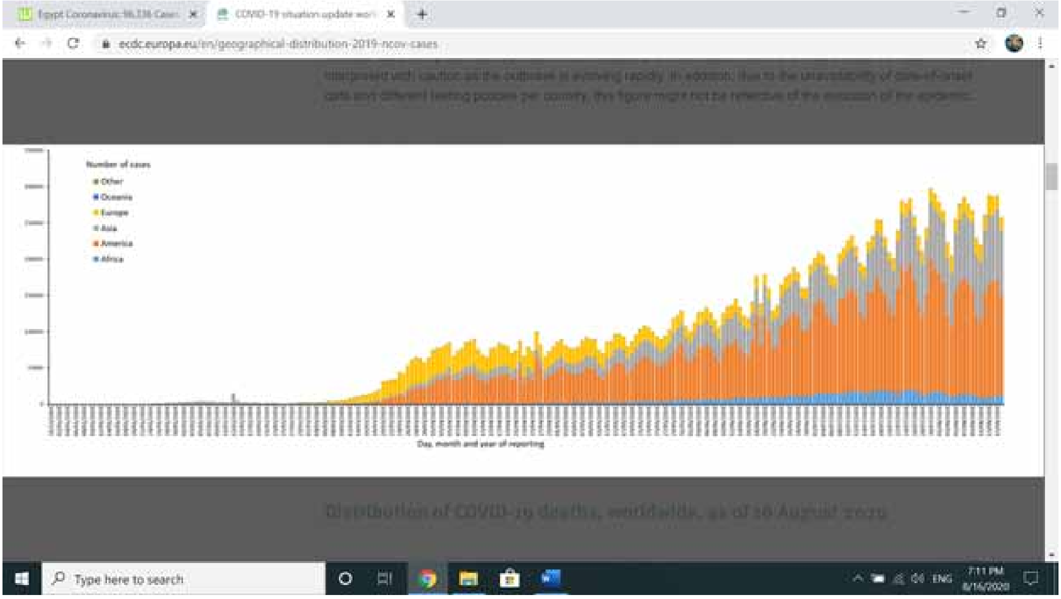
As COVID-19 is declared as an international epidemic in mid of March 2020 and more than 125000 confirmed cases have been recorded around the world, so day level forecasting about COVID -19 spread is crucial to measure the behavior of this new virus globally (WHO, 2020).

Corona virus is spreading around the world, causing panic to all men, for regional distribution of COVID-19 cases worldwide see Figure 1. It causes serious breathing-related symptoms particularly for elderly people and those suffering from chronic illnesses. For the past 6 months, diagnosing Corona virus disease is too complicated as it takes a long time to get results for COVID-19 tests based on the signs and symptoms. There is still no approved antiviral drug for treating COVID-19 until now.

DOI: 10.4018/IJHISI.294115

*Corresponding Author

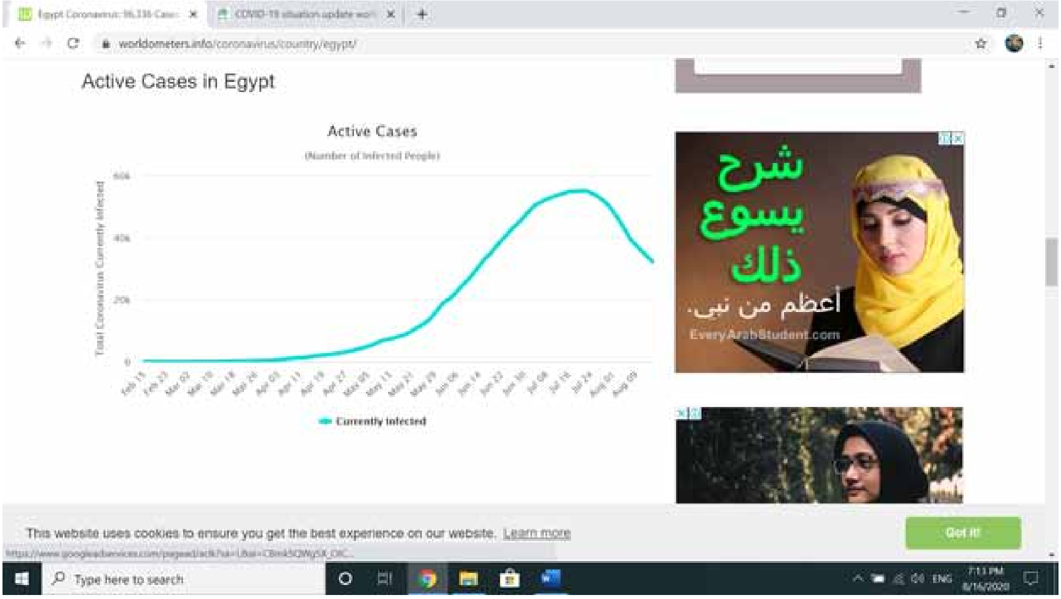
Figure 1. Geographic distribution of COVID-19 cases worldwide, as of 16 August 2020 (ecdc)


Egypt Coronavirus: 96,336 Cases
COVID-19 situation update worl
ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases
Number of cases
Other
Oceania
Europe
Asia
America
Africa
Day, month and year of reporting
Distribution of COVID-19 deaths, worldwide, as of 16 August 2020
Type here to search
ENG
7:11 PM
8/16/2020

From figure 1, the total numbers of infected cases of corona virus over the world is constantly increased during the past six months.

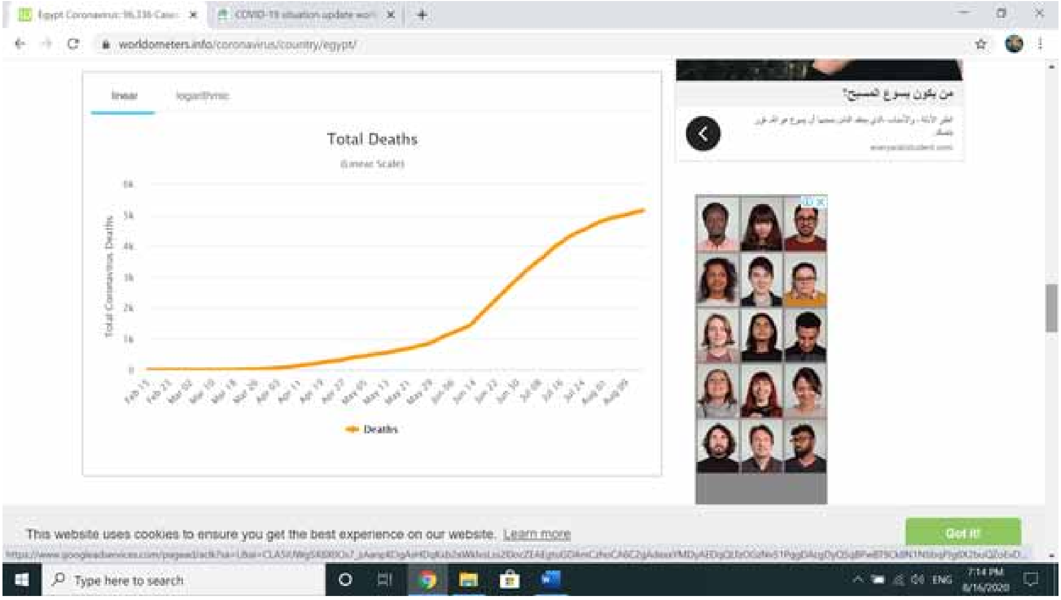
In Egypt, according to the Egyptian ministry of health, the infected cases of Corona virus are shown in figure 2. And the death cases of Corona virus are shown in figure 3.

Figure 2. Total Corona virus Cases in Egypt, as of 16 August 2020 (worldometers)



This website uses cookies to ensure you get the best experience on our website. Learn more

Figure 3. Total Corona virus death Cases in Egypt, as of 16 August 2020 (worldometers)



This website uses cookies to ensure you get the best experience on our website. Learn more

From figures 2 and 3, we notice the total numbers of corona virus in Egypt start with low infected and death cases during the first three months from mid-February to mid-May then total number of infected and death cases increase quickly with high numbers from mid-May to late July then total numbers start to decrease during August.

Corona virus causes a real health crisis around the world, beside its social, economic and political effects, and to help in solving this problem and avoiding infection and planning the healthcare system for possible potential up-comings, there is an urgent need to construct forecasting model providing future forecasts of the possible number of daily cases and daily deaths. In some case The accuracy of traditional forecasting is not high because it depends primarily on the availability of data but, in disease outbreaks such as current corona virus outbreak, there are no data at all at first and then limited as time goes by, there are concerns that the data may not be reliable.

Despite the inaccuracies associated with medical forecasts, forecasting is still valuable in enabling us to better understand the current situation and to plan ahead. This paper presents a computational competent and realistic forecasting model for infected and death cases in Egypt, which can help the policy makers and the medical system in preparation for the new patients. Also help to face the general population fears about the impact of corona virus and increase the public awareness.

Various models for COVID-19 infected and death cases predictions have been implemented, since the COVID-19 outbreak in Wuhan City in December of 2019. The prediction models have shown a wide range of diversity due to the non-identifiable in model calibrations using the confirmed-case data is the main cause for this diversity. As governments in any countries needs to know the upcoming infected as to help the public health agencies making decisions.

LITERATURE REVIEW

SVM algorithm has a role in all major industries such as banking, healthcare, transport and media. In these days, healthcare subject is advancing quickly with information and difficulties in patient outcomes. As healthcare is affecting not only the developed countries but also European countries economically, the availability of coronavirus medical data and the forecasting of health and economic situation are needed. It is not possible by using conventional techniques to analyze this situation and solve the problem. SVM works relatively well when there is a clear margin of separation between classes. It is more effective in high dimensional spaces. SVM is effective in cases where the number of dimensions is greater than the number of samples. Therefore, SVM technique is coming up to solve these issues as it is suitable for available data volume, information type, and outcomes related to requirements (Ichikawa D, et al., 2016).

The transmission mechanism and major factors influencing the COVID-19 spread, such as the number of basic regenerations, the incubation time and the average number of cure days is examined in (Li et al., 2020). The research forecasted the evolution of current epidemic data in South Korea, Italy, and Iran, indicating the recrudescence as to help these countries to control this epidemic and give some references for future research. Other studies are needed to forecast COVID-19 cases in other countries.

A study used data on the number of cases exported from Wuhan internationally to forecast the national spread of COVID-19, accounting for the effect of the metropolitan-wide quarantine of Wuhan and surrounding cities using Markov Chain Monte Carlo method. The study assumed that if the transmissibility of COVID-19 were similar everywhere, therefore, that epidemics are already growing exponentially in multiple major cities of China (Wu et al.,2020). The assuming that the transmission process of COVID-19 is analogous in all countries cannot be accepted, so this makes the forecasting of COVID-19 cases evolution is important.

Logistic model, Bertalanffy model and Gompertz model were applied to fit and analyze the situation of COVID-19. The forecasting results of the three mathematical models are different for different parameters and regions. From their results the Logistic model's fitting effect may be the best of the three models. Based on the three models, the total number expected to be infected in Wuhan is fifty thousand people in three months (jia et al., 2020).

A study to examine Auto-Regressive Integrated Moving Average (ARIMA) model for the prediction of confirmed COVID-19 infection cases in India was applied (Tandon et.al., 2020). Another study was developed for forecasting future COVID-19 cases in India to identify the best predictive models for confirmed cases on a regular basis in countries with a large number of reported cases worldwide and second, to forecast confirmed cases using such models in order to be more equipped for health systems (Dehesh et al., 2020). The forecasting results showed that the confirmed cases are expected to be increased. As needed data is not available in the present circumstance, the ARIMA predicting model is valuable in predicting future cases if the manner of virus spread didn't change abnormally.

A prediction model is developed to predict the continuation of the COVID-19 assuming that the data used is reliable. The forecasting results declare a continuing increase in the confirmed COVID-19 cases with considerable uncertainty. The results present the timeline of a live forecasting and provides objective forecasts for the confirmed cases of COVID-19 (Petropoulos & Makridakis,2020).

Another study used the SIR and SEIR models for Covid-19 cases prediction, the study showed that SIR model performs much better than an SEIR model in representing the information contained in the confirmed-case data. The results indicated that predictions using complex models may not be accurate compared to using a simpler model (Roda, W., C., 2020).

Several techniques have been implemented to predict infected and death Covid-19 cases but none of the techniques are able to provide an accurate result till now. Therefore, this paper presents SVM technique which is related with learning computations to investigate the data utilized for regression analysis and classification to Covid-19 cases prediction.

PSEUDO CODE OF THE PROPOSED MODEL

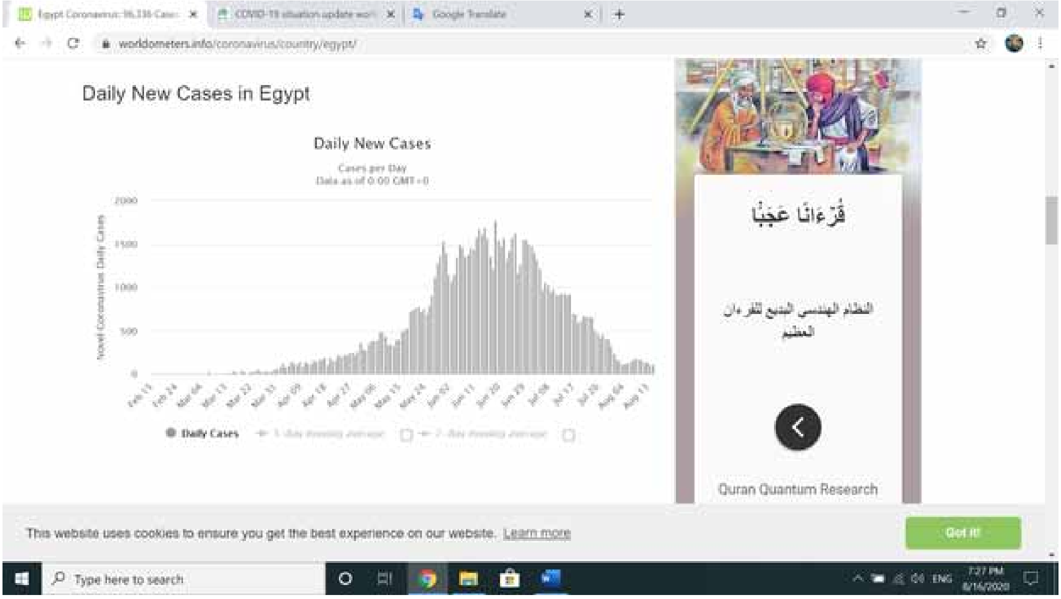
```
1- Collecting Covid 19 data.
```

2- Data contains daily infected cases and daily death cases.
3- Using Weka time serious forecasting packet.
4- Applying simple time series forecasting algorithms:
  a. Support vector machine (SVM) algorithm.
  b. Linear regression algorithm.
  c. Random Forest algorithm.
5- Forecasting confirmed and death cases of COVID-19.
6- For each month:
  a. Produce 30-days-ahead point forecasts
  b. Update forecasts every month.
7- Measuring the accuracy of forecasting results using:
  a. Mean absolute error (MAE)
  b. Root Mean Square Error (RMSE)
  c. Mean directional accuracy (MDA)

ANALYSIS AND FORECASTING

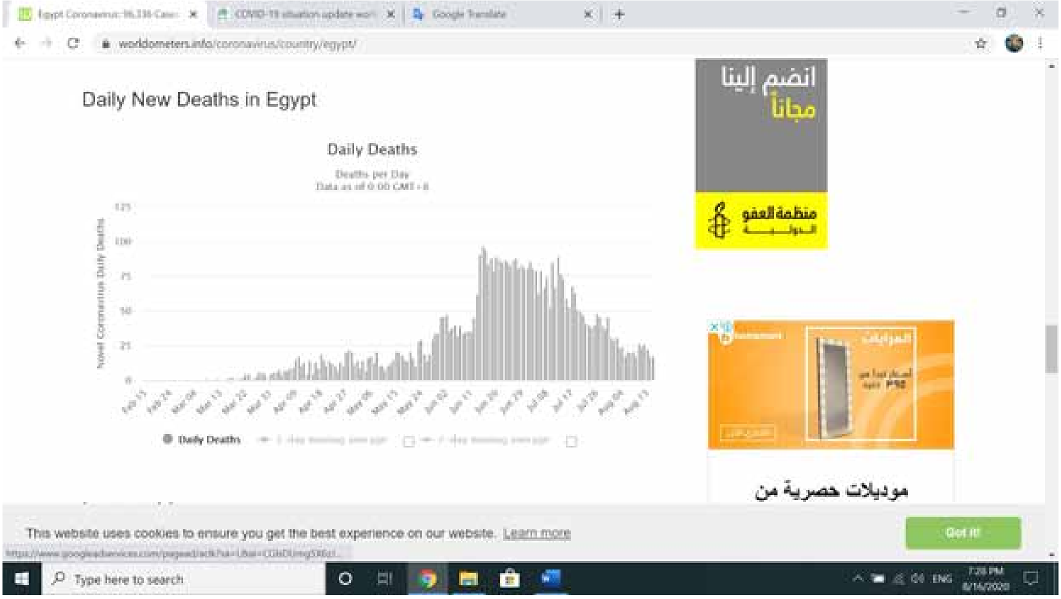
We focus on the daily figures with the two main variables of interest: confirmed cases, deaths. These data were provided by *Egyptian Ministry of Health* (https://www.care.gov.eg/EgyptCare/Index.aspxlastaccessedon10/05/2020) as presented in figures 4 and 5. Data as features is described as daily infected cases and daily death cases. To forecast confirmed and death cases of COVID-19, we used Weka time serious forecasting packet to adopt simple time series forecasting approach using support vector machine (SVM) algorithm. We produce 30-days-ahead point forecasts and prediction intervals and update forecasts every month.

Figure 4. Daily new cases of corona in Egypt as of 16 August 2020 (worldometers)



This website uses cookies to ensure you get the best experience on our website. Learn more

Figure 5. Daily deaths cases of corona in Egypt as of 16 August 2020 (worldometers)



This website uses cookies to ensure you get the best experience on our website. Learn more

Got it!

From figures 4 and 5, we notice that daily numbers of infected and death cases in Egypt start with low numbers during the first four months from February to June, then daily numbers start increase quickly during late June to late July then daily numbers start to decrease again during *August*.

SVM

One of the most important machine learning algorithms is SVM (Support vector Machine) which was presented in 1995 depending on statistical theory. The advantage of this method appears in small sample, nonlinear and high dimensional pattern recognition (Cortes & Vapnik, 1995). SVM fundamental is to handle complicated data classification by finding answer to the optimization problem (Haung, 2018). SVM main objective is to build a decision border called hyper-plane between two classes that allows labels prediction from one or more feature vectors. The closest data points from each of the classes are called support vectors. Given a labeled training dataset (Haung, 2018):

$(x_1, y_1), ..., (x_n, y_n)$, xi Î R^d and y_i Î (−1, +1)

Where x_i is a feature vector representation and y_i the class label (negative or positive) of training compound $_i$.

SVM uses a linear model to apply nonlinear class boundaries by mapping input vectors x to high-dimensional feature space by certain nonlinear ones. In the original space, a linear model built into the new space may represent a nonlinear boundary of decision. An optimal separating hyper-plane is constructed in the new space. SVM is thus known as the algorithm which finds a special form of the linear model, the maximum hyper-plane margin. The maximum hyper-plane margin provides the maximum distinction between the classes of decisions. The training examples which are nearest to the hyper-plane of the maximum margin are called support vectors. All training examples are meaningless for determining the boundaries of binary classes (Kyoung, 2003).

Linear Regression

Linear regression makes the assumption of linearity. This assumption makes the model easy to interpret but is often not flexible enough for prediction (Schonlau & Zou, 2020). Linear regression is explicit upright approach to predict quantitative response Y based on a sin regression predictor variable X. It assumes that there is approximately a linear relationship between X and Y (James et al., 2013). It can be expressed as Y » β0 + β_1X.

Random Forest

Random Forest is a mathematically effective method that can implemented over large datasets. It has been used in many recent research projects and real-world applications in diverse domains (Belgiu & Drăguţ, 2016). Also, Random Forest is a good approach for handling missing data including interactions and nonlinearity. There are many Random Forest algorithms to handle large and diverse data sets. Random Forest imputation is enhanced with increasing correlation. Performance was accepted under moderate to and high data missing (Tang & Ishwaran, 2017).

Random decision forests can handle nonlinearities for medium to large datasets which makes this method better than linear regression in prediction. Linear regression is not suitable when the number of independent variables is greater than the number of observations, while random forest can handle this case as not all predictor variables are used at once (Schonlau & Zou, 2020).

Evaluation

Mean absolute error (MAE) “is a measure of errors between paired observations expressing the same phenomenon”. It measures accuracy of continuous variables MAE is calculated as (Chai & Draxler, 2014):

$$\frac{1}{n}\sum_{\mathrm{i=1}}^{\mathrm{n}}\left|\mathrm{ei}\right| \tag{1}$$

Root Mean Square Error (RMSE) “is the standard deviation of the residuals (prediction errors)” (Petropoulos & Makridakis, 2020):

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}e_{\mathrm{i}}^{2}} \tag{2}$$

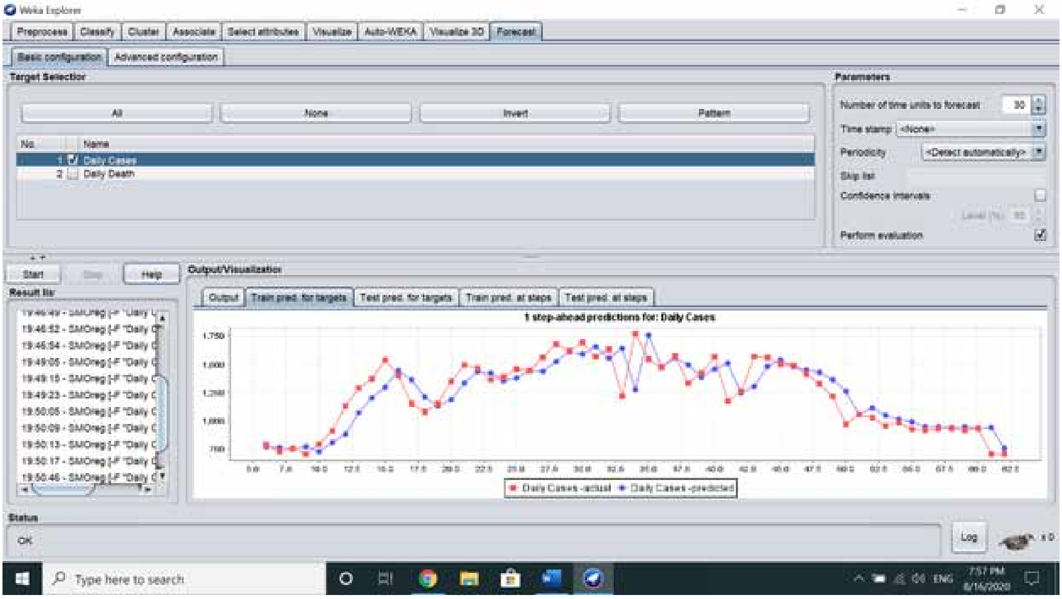
To simplify, we assume that we already have n samples of model errors n calculated as (e_i, i = 1,2,..., n).

Mean directional accuracy (MDA) “is a measure of prediction accuracy of a forecasting” It compares the forecast direction to the actual realized direction. It is defined by the following formula (Pesaran & Timmermann, 2001):

$$\frac{1}{\mathrm{N}}\sum_{t} 1_{\mathrm{Sign}}\,(\mathrm{A}_t\text{-}\mathrm{A}_{t\text{-}1}) == \mathrm{sign}\,(\mathrm{F}_t\text{-}\mathrm{A}_{t\text{-}1}) \tag{3}$$

Where A_t is the actual value at time t and F_t is the forecast value at time t. Variable N represents number of forecasting points.

Figure 6. A comparative time series plot for actual infected cases and forecasted COVID-19 cases from 15 March 2020 to 16 August 2020 for training data.



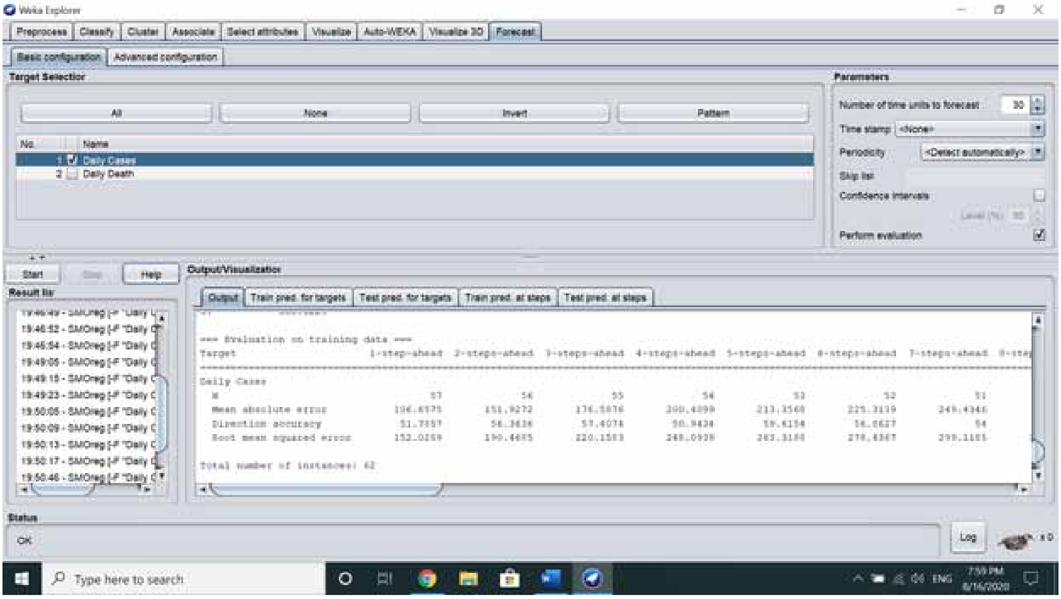
Parameters
Number of time units to forecast 30
Time stamp <None>
Periodicity <Detect automatically>
Skip list
Confidence intervals □
Level (%) 95
Perform evaluation ☑

No.	Name
1	☑ Daily Cases
2	□ Daily Death

Result list
19:46:52 - SMOreg [-F "Daily C
19:46:54 - SMOreg [-F "Daily C
19:49:05 - SMOreg [-F "Daily C
19:49:15 - SMOreg [-F "Daily C
19:49:23 - SMOreg [-F "Daily C
19:50:05 - SMOreg [-F "Daily C
19:50:09 - SMOreg [-F "Daily C
19:50:13 - SMOreg [-F "Daily C
19:50:17 - SMOreg [-F "Daily C
19:50:46 - SMOreg [-F "Daily C

Status
OK

Figure 7. Evaluation metrics for forecasting of infected cases for training data



Parameters
Number of time units to forecast 30
Time stamp <None>
Periodicity <Detect automatically>
Skip list
Confidence intervals □
Level (%) 95
Perform evaluation ☑

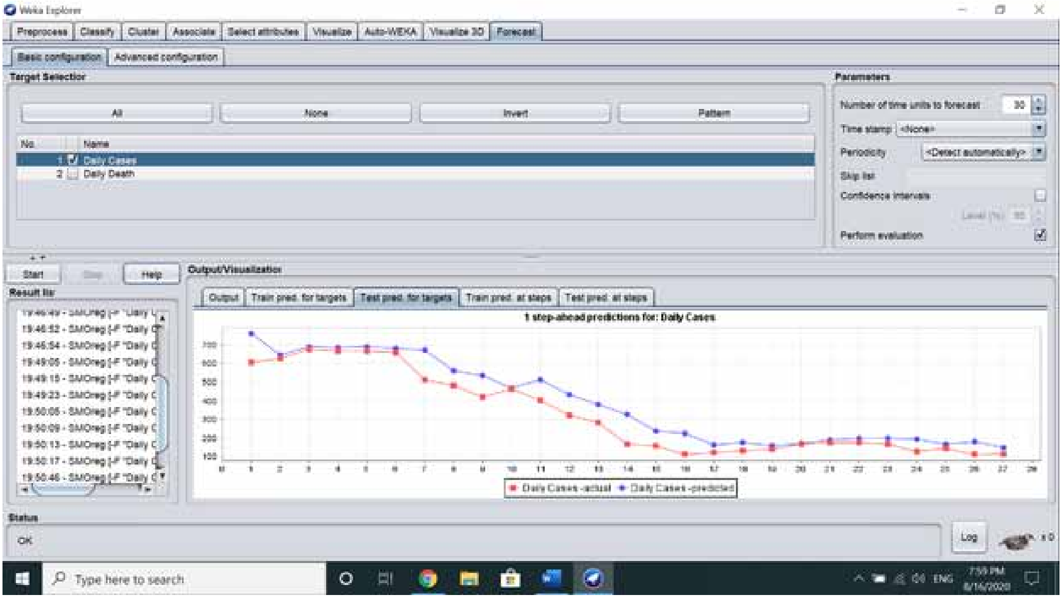
No.	Name
1	☑ Daily Cases
2	□ Daily Death

Result list
19:46:52 - SMOreg [-F "Daily C
19:46:54 - SMOreg [-F "Daily C
19:49:05 - SMOreg [-F "Daily C
19:49:15 - SMOreg [-F "Daily C
19:49:23 - SMOreg [-F "Daily C
19:50:05 - SMOreg [-F "Daily C
19:50:09 - SMOreg [-F "Daily C
19:50:13 - SMOreg [-F "Daily C
19:50:17 - SMOreg [-F "Daily C
19:50:46 - SMOreg [-F "Daily C

```
=== Evaluation on training data ===
Target                         1-step-ahead  2-steps-ahead  3-steps-ahead  4-steps-ahead  5-steps-ahead  6-steps-ahead  7-steps-ahead  8-step
=========================================================================================================================================
Daily Cases
  N                                      57             56             55             54             53             52             51
  Mean absolute error              196.6575       151.9272       176.5876       200.4099       213.3568       225.3139       249.4346
  Direction accuracy                51.7857        56.3636        57.4074        50.9434        59.6154        56.8627             54
  Root mean squared error          152.0259       190.4685       220.1583       248.0939       263.3188       278.4367       299.1185

Total number of instances: 62
```

Status
OK

Figure 8. A comparative time series plot for actual infected cases and forecasted COVID-19 cases from 15 March 2020 to 16 August 2020 for testing data.



No.	Name
1 ☑	Daily Cases
2 □	Daily Death

Parameters

Number of time units to forecast 30

Time stamp <None>

Periodicity <Detect automatically>

Skip list

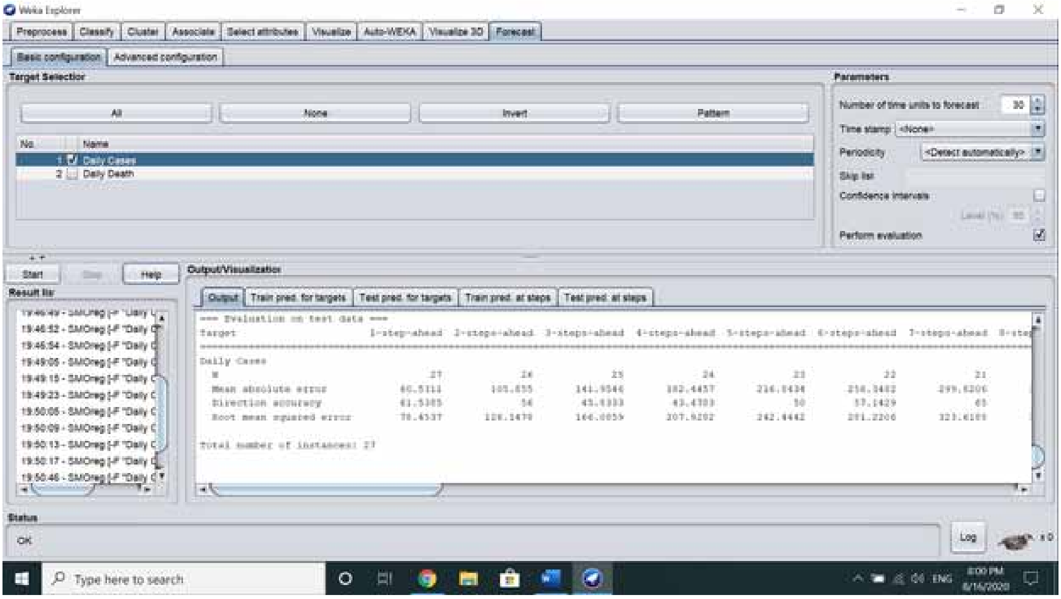
Confidence intervals □

Level (%) 95

Perform evaluation ☑

Status: OK

Figure 9. Evaluation metrics for forecasting of infected cases for testing data



No.	Name
1 ☑	Daily Cases
2 □	Daily Death

Parameters

Number of time units to forecast 30

Time stamp <None>

Periodicity <Detect automatically>

Skip list

Confidence intervals □

Level (%) 95

Perform evaluation ☑

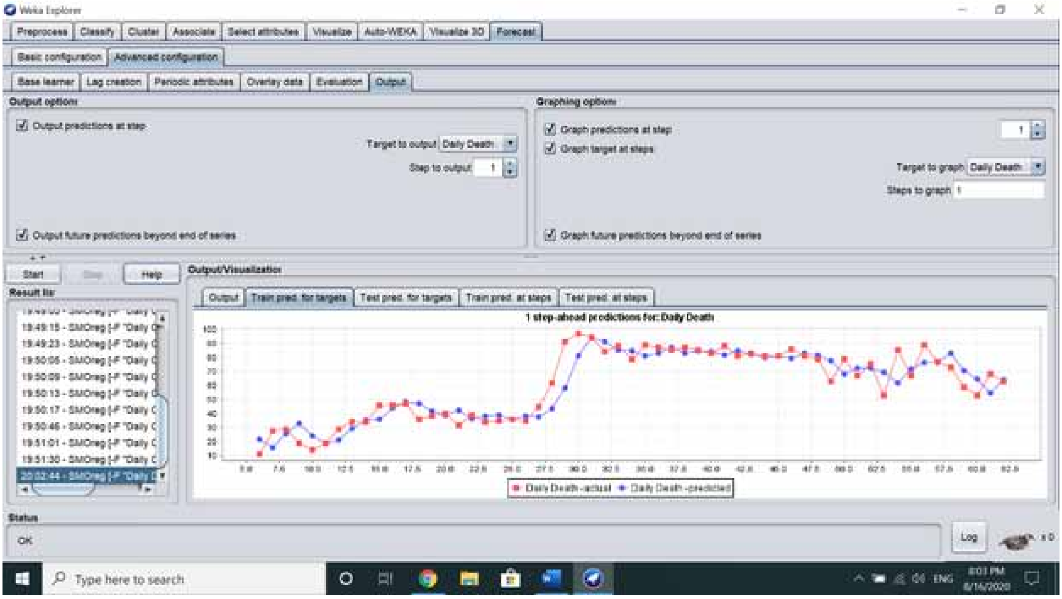
```
=== Evaluation on test data ===
Target                    1-step-ahead  2-steps-ahead  3-steps-ahead  4-steps-ahead  5-steps-ahead  6-steps-ahead  7-steps-ahead  8-ste
=========================================================================================================================================
Daily Cases
  N                                27             26             25             24             23             22             21
  Mean absolute error         80.5111        105.855       141.9546       182.4457       216.0434       258.3402       299.6206
  Direction accuracy          61.5385             56        45.8333        43.4783             50        57.1429             65
  Root mean squared error     98.4537       128.1479       166.0059       207.9202       242.4442       271.2206       323.6185

Total number of instances: 27
```

Status: OK

DISCUSSION AND RESULTS

For comparing the actual and forecasted infected COVID-19 cases for training data, a time series graph is plotted starting from 15 March 2020 till *16 August* 2020. The plot is represented by figure 6. The high similarity of forecasted data with actual data and the mean absolute error values and accuracy values as shown in figure 7 reveal very acceptable precision of the model in forecasting of infected corona cases.

Figure 10. A comparative time series plot for actual death cases and forecasted COVID-19 cases from 15 March 2020 to 16 August 2020 for training data.



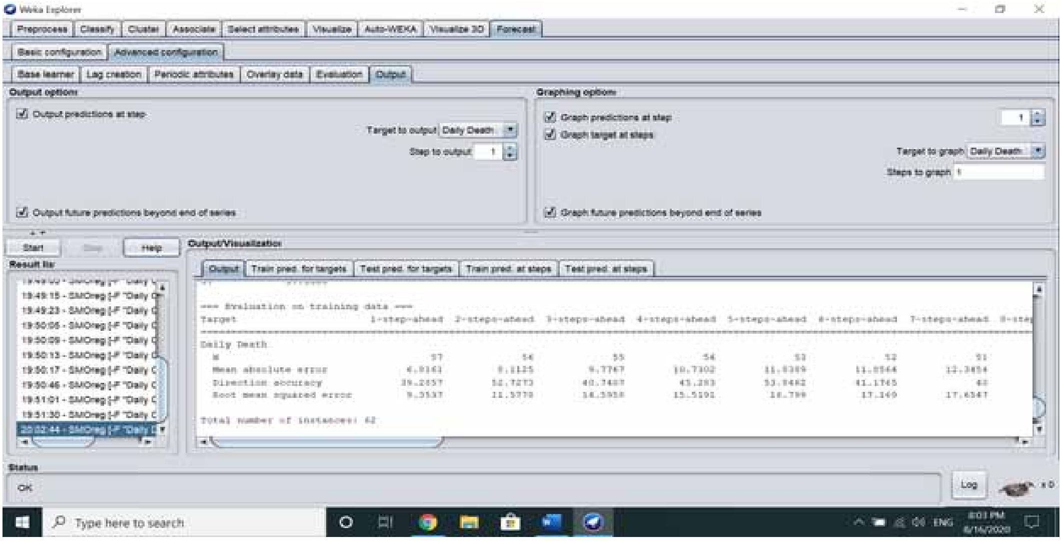
Output options
☑ Output predictions at step
Target to output Daily Death
Step to output 1
☑ Output future predictions beyond end of series

Graphing options
☑ Graph predictions at step 1
☑ Graph target at steps
Target to graph Daily Death
Steps to graph 1
☑ Graph future predictions beyond end of series

Result list
19:49:15 - SMOreg [-F "Daily D
19:49:23 - SMOreg [-F "Daily D
19:50:05 - SMOreg [-F "Daily D
19:50:09 - SMOreg [-F "Daily D
19:50:13 - SMOreg [-F "Daily D
19:50:17 - SMOreg [-F "Daily D
19:50:46 - SMOreg [-F "Daily D
19:51:01 - SMOreg [-F "Daily D
19:51:30 - SMOreg [-F "Daily D
20:02:44 - SMOreg [-F "Daily D

Status
OK

Figure 11. Evaluation metrics for forecasting of death cases for training data



Output options
☑ Output predictions at step
Target to output Daily Death
Step to output 1
☑ Output future predictions beyond end of series

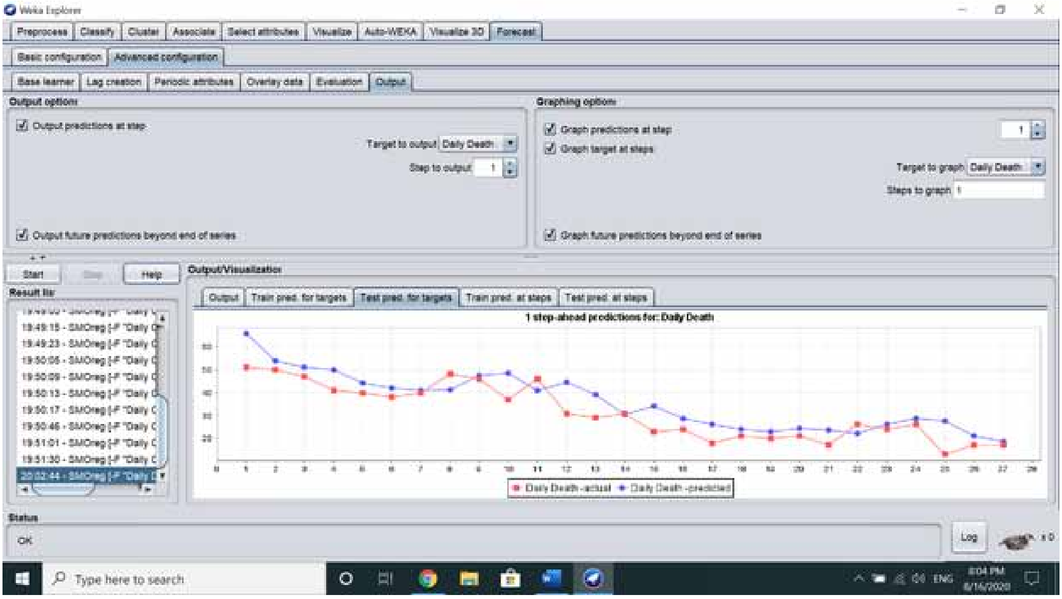
Graphing options
☑ Graph predictions at step 1
☑ Graph target at steps
Target to graph Daily Death
Steps to graph 1
☑ Graph future predictions beyond end of series

Result list
19:49:15 - SMOreg [-F "Daily D
19:49:23 - SMOreg [-F "Daily D
19:50:05 - SMOreg [-F "Daily D
19:50:09 - SMOreg [-F "Daily D
19:50:13 - SMOreg [-F "Daily D
19:50:17 - SMOreg [-F "Daily D
19:50:46 - SMOreg [-F "Daily D
19:51:01 - SMOreg [-F "Daily D
19:51:30 - SMOreg [-F "Daily D
20:02:44 - SMOreg [-F "Daily D

```
=== Evaluation on training data ===
Target                     1-step-ahead  2-steps-ahead  3-steps-ahead  4-steps-ahead  5-steps-ahead  6-steps-ahead  7-steps-ahead  8-ste
========================================================================================================================================
Daily Death
  N                                  57             56             55             54             53             52             51
  Mean absolute error            6.8161         8.1125         9.7767        10.7302        11.8389        11.8564        12.3854
  Direction accuracy            39.2857        52.7273        40.7407         45.283        53.8462        41.1765             40
  Root mean squared error        9.3537        11.5778        14.5958        15.5191         16.799         17.169        17.6547

Total number of instances: 62
```

Status
OK

Figure 12. A comparative time series plot for actual infected cases and forecasted COVID-19 cases from 15 March 2020 to 16 August 2020 for testing data.

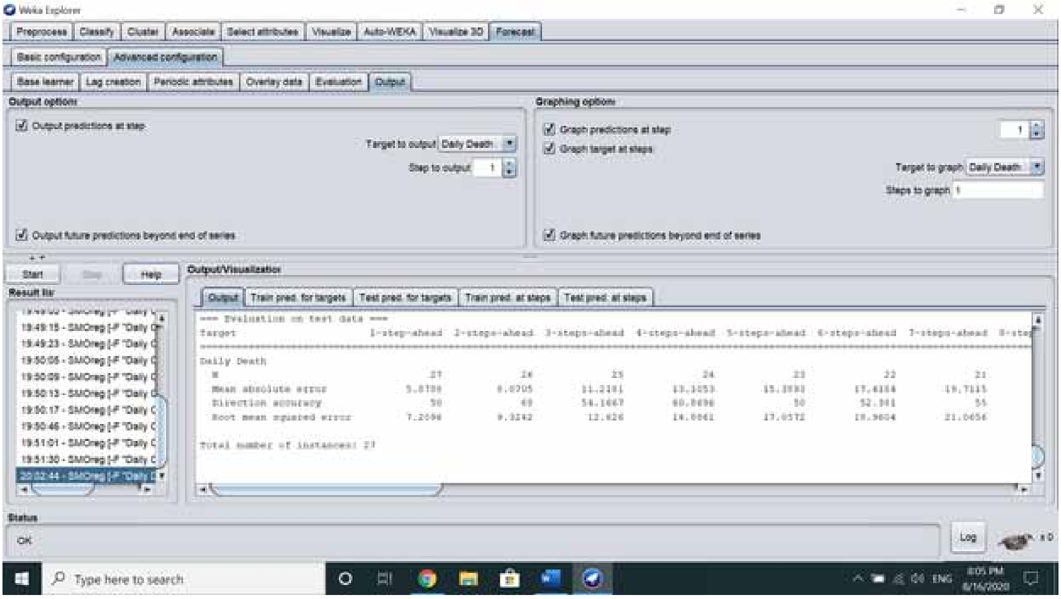


Output options
☑ Output predictions at step
Target to output Daily Death
Step to output 1
☑ Output future predictions beyond end of series

Graphing options
☑ Graph predictions at step 1
☑ Graph target at steps
Target to graph Daily Death
Steps to graph 1
☑ Graph future predictions beyond end of series

Status
OK

Figure 13. Evaluation metrics for forecasting of death cases on testing data



Output options
☑ Output predictions at step
Target to output Daily Death
Step to output 1
☑ Output future predictions beyond end of series

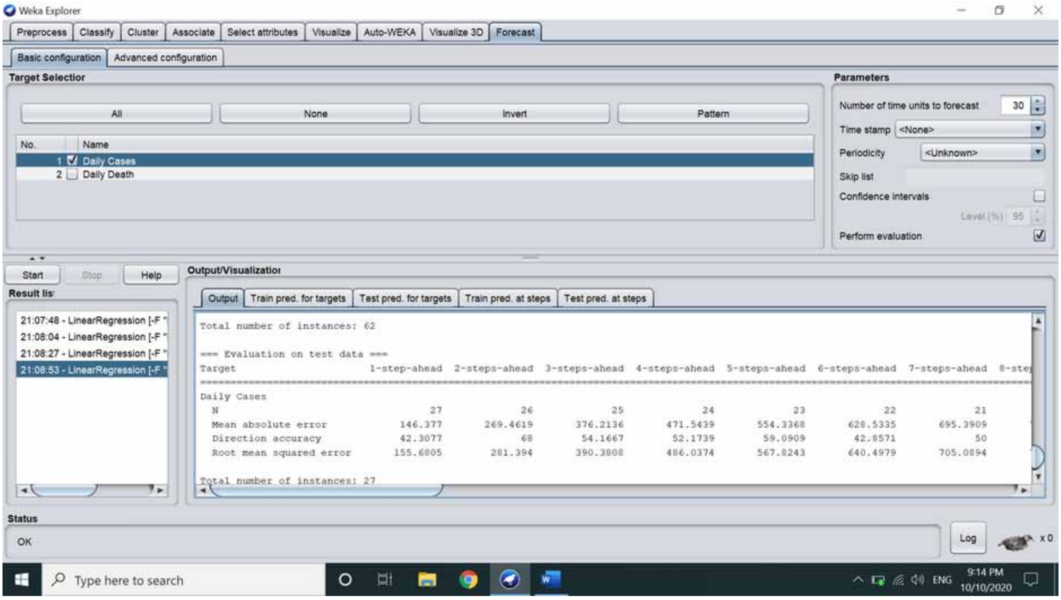
Graphing options
☑ Graph predictions at step 1
☑ Graph target at steps
Target to graph Daily Death
Steps to graph 1
☑ Graph future predictions beyond end of series

```
=== Evaluation on test data ===
Target                         1-step-ahead  2-steps-ahead  3-steps-ahead  4-steps-ahead  5-steps-ahead  6-steps-ahead  7-steps-ahead  8-step
================================================================================================================================================
Daily Death
  N                                      27             26             25             24             23             22             21
  Mean absolute error                5.8788         8.0705        11.2181        13.1053        15.3593        17.4384        19.7115
  Direction accuracy                     50             60        54.1667        60.8696             50         52.381             55
  Root mean squared error            7.2098         9.3242         12.626        14.8861        17.0572        18.9604        21.0656

Total number of instances: 27
```

Status
OK

For comparing the actual and forecasted infected COVID-19 cases for testing data, a time series graph is plotted starting from 15 March 2020 till *16 August* 2020. The plot is represented by figure 8. The high similarity of forecasted data with actual data and the mean absolute error values and accuracy values as shown in figure 9 reveal the high precision of the model in forecasting of infected corona cases.

Figure 14. Evaluation metrics for forecasting of infected cases on testing data using linear regression



Target Selection

No.	Name
1	☑ Daily Cases
2	☐ Daily Death

Parameters

Number of time units to forecast 30
Time stamp <None>
Periodicity <Unknown>
Skip list
Confidence intervals ☐
Level (%) 95
Perform evaluation ☑

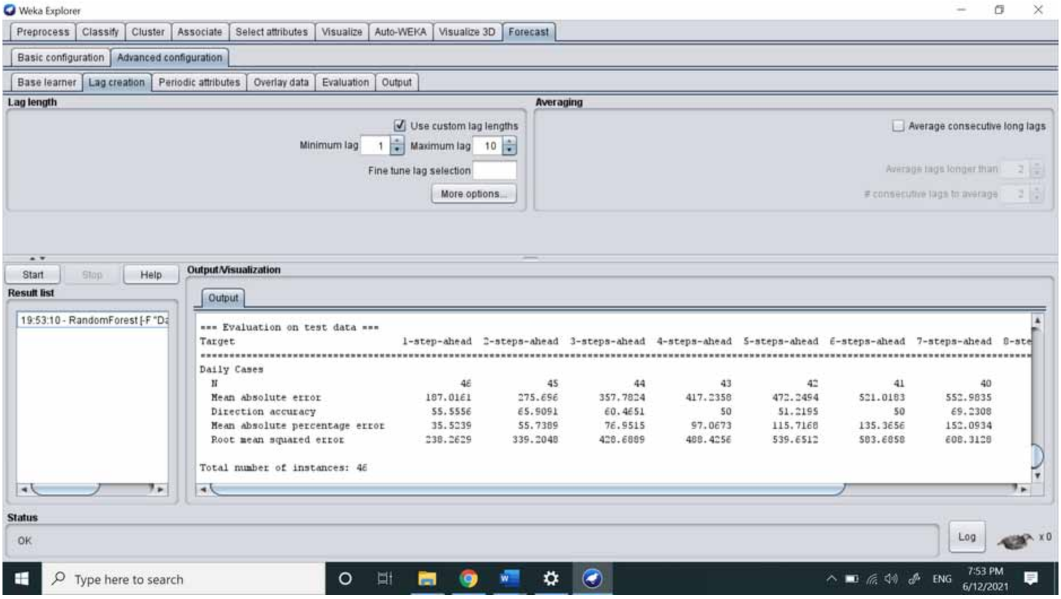
Result list

21:07:48 - LinearRegression [-F "
21:08:04 - LinearRegression [-F "
21:08:27 - LinearRegression [-F "
21:08:53 - LinearRegression [-F "

```
Total number of instances: 62

=== Evaluation on test data ===
Target                         1-step-ahead  2-steps-ahead  3-steps-ahead  4-steps-ahead  5-steps-ahead  6-steps-ahead  7-steps-ahead  8-ste
==========================================================================================================================================
Daily Cases
  N                                     27             26             25             24             23             22             21
  Mean absolute error              146.377       269.4619       376.2136       471.5439       554.3368       628.5335       695.3909
  Direction accuracy               42.3077             68        54.1667        52.1739        59.0909        42.8571             50
  Root mean squared error         155.6805        281.394       390.3808       486.0374       567.8243       640.4979       705.0894

Total number of instances: 27
```

Status
OK

Figure 15. Evaluation metrics for forecasting of infected cases on testing data using random forest



Lag length

☑ Use custom lag lengths
Minimum lag 1
Maximum lag 10
Fine tune lag selection

Averaging

☐ Average consecutive long lags
Average lags longer than 2
# consecutive lags to average 2

Result list

19:53:10 - RandomForest [-F "Da

```
=== Evaluation on test data ===
Target                              1-step-ahead  2-steps-ahead  3-steps-ahead  4-steps-ahead  5-steps-ahead  6-steps-ahead  7-steps-ahead  8-ste
===============================================================================================================================================
Daily Cases
  N                                          46             45             44             43             42             41             40
  Mean absolute error                  187.0161        275.696       357.7824       417.2358       472.2494       521.0183       552.9835
  Direction accuracy                    55.5556        65.9091        60.4651             50        51.2195             50        69.2308
  Mean absolute percentage error        35.5239        55.7389        76.9515        97.0673       115.7168       135.3656       152.0934
  Root mean squared error              238.2629       339.2048       428.6889       488.4256       539.6512       583.6858       608.3128

Total number of instances: 46
```

Status
OK

For comparing the actual and forecasted death COVID-19 cases for training data, a time series graph is plotted starting from 15 March 2020 till *16 August* 2020. The plot is represented by figure 10. The high similarity of forecasted data with actual data and the mean absolute error values and accuracy as shown in figure 11 values reveal the high precision of this model in forecasting of death corona cases.

For comparing the actual and forecasted death COVID-19 cases for testing data, a time series graph is plotted starting from 15 March 2020 till *16 August* 2020. The plot is represented by figure 12. The high similarity of forecasted data with actual data and the mean absolute error values and accuracy as shown in figure 13 values reveal the high precision of this model in forecasting of death corona cases.

The forecasting results for Corona Virus infected and death cases using SVM are more accurate than linear regression forecasting results according to Mean absolute error and accuracy metrics. To show that, we select to present an example of forecasting results of Corona Virus infected cases for testing data using linear regression in figure 14 compared to forecasting results for infected cases using SVM (see figure 9).

The forecasting results for Corona Virus infected and death cases using SVM are more accurate than random forest forecasting results according to Mean absolute error and accuracy metrics. To show that, we select to present an example of forecasting results of Corona Virus infected cases for testing data using random forest in figure 15 compared to forecasting results for infected cases using SVM (see figure 9).

We applied two more algorithms in the prediction of coronavirus infected and death cases; Linear Regression Model and Random Forest forecasting to compare their results with SVM forecasting results. SVM model achieved more accurate forecasting results with less MAE. Next table shows MAE and accuracy metrics for the three algorithms. SVM prediction model achieved the least MAE with 60 and highest accuracy with almost 62%. Linear Regression Model achieved high MAE with 146 and least accuracy with almost 42%, While Random Forest Prediction Model achieved highest MAE with 187 and low accuracy with almost 56%.

Table 1. Comparing the proposed model with the state-of-the-art

	SVM Prediction Model	Linear Regression Model	Random Forest Prediction Model
Mean absolute error	60.531	146.337	187.016
accuracy	61.539%	42.308%	55.556%

The research question is what are the expected number of infected and death cases. From the experimental results, it can be concluded that the prediction results of infected and death cases of Corona virus based on the proposed model has accurate results. The mean error was high for the infected cases and was low for the death cases. Also, the mean error increases with the more prediction steps. The prediction model expects the numbers of infected and death cases will increase again after one month, So it is recommended that more medical equipment and efforts are needed to face the crisis of Corona virus and effective economic strategies are needed to deal with the effects of expected high spreading of the Corona virus disease. Also, more efforts shall be directed the increase the public awareness with this disease.

CONCLUSION

COVID-19 is a new virus spread quickly declared as an international epidemic. This research recommends that all world countries must mandate restrictions on public gatherings to decrease the infected cases. From the experimental results, the forecasting results in death and infected cases of corona virus in Egypt are accurate and acceptable compared to other algorithms forecasting results. The forecasting model should be playing a big role in the decision-making process and help the medical system in preparation for the new patients. Also help to face the general population fears about the impact of corona virus and increase the public awareness apart from its effect on global supply chains and the economy.

REFERENCES

Battineni, G., Chintalapudi, N., & Amenta, F. (2019). Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Informatics in Medicine Unlocked*, *16*, 100200. doi:10.1016/j.imu.2019.100200

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. doi:10.1016/j.isprsjprs.2016.01.011

Chai, T., & Draxler, R. (2014). *Root mean square error (RMSE) or mean absolute error (MAE). eosci*. Model Dev.

Coronavirus. (n.d.). https://www.worldometers.info/coronavirus/

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. doi:10.1007/BF00994018

Dehesh, T., Mardani-Fard, H. A., & Dehesh, P. (2020). *Forecasting of COVID-19 Confirmed Cases in Different Countries with ARIMA Models.* Populations and Evolution Egyptian Ministry of Health. https://www.care.gov.eg/EgyptCare/Index.aspx

Gholami, R., & Fakhari, N. (2017). *Learn more about support vector machine support vector Machine: Principles, parameters, and applications quantitative structure-activity relationship (QSAR): Modeling approaches to biological applications technical aspects of brain rhythms and sp*. Academic Press.

Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics & Proteomics*, *15*(1), 41–51. PMID:29275361

Ichikawa, D., Saito, T., Ujita, W., & Oyama, H. (2016). How can machine-learning methods assist in virtual screening for hyperuricemia? A healthcare machine-learning approach. *Journal of Biomedical Informatics*, *64*, 20–24. doi:10.1016/j.jbi.2016.09.012 PMID:27658886

Islam, M. M., Iqbal, H., Haque, M. R., & Hasan, M. K. (2017, December). Prediction of breast cancer using support vector machine and K-Nearest neighbors. In *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 226-229). IEEE.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Linear regression. In *An introduction to statistical learning* (pp. 59–126). Springer. doi:10.1007/978-1-4614-7138-7_3

Jia, L., Li, K., Jiang, Y., & Guo, X. (2020). Prediction and analysis of Coronavirus Disease 2019. Stanford University.

Kyoung, J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1–2), 307–319.

Li, L., Yang, Z., Dang, Z., Meng, C., Huang, J., Meng, H., Wang, D., Chen, G., Zhang, J., Peng, H., & Shao, Y. (2020). Propagation analysis and prediction of the COVID-19. *Infectious Disease Modelling*, *5*, 282–292. doi:10.1016/j.idm.2020.03.002 PMID:32292868

Pesaran, M. H., & Timmermann, A. (2004). A. How costly is it to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting*, *20*(3), 411–425. doi:10.1016/S0169-2070(03)00068-2

Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PLOS ONE Journal*.

Roda, W. C., Varughese, M. B., Han, D., & Li, M. Y. (2020). Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling*, *5*, 271–281. doi:10.1016/j.idm.2020.03.001 PMID:32289100

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, *20*(1), 3–29. doi:10.1177/1536867X20909688

Tandon, H., Ranjan, P., Chakraborty, T., & Suhag, V. (2020). *Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future*. Populations and Evolution.

Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *10*(6), 363–377. doi:10.1002/sam.11348 PMID:29403567

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modeling study. *Lancet*, *395*(10225), 689–697. doi:10.1016/S0140-6736(20)30260-9 PMID:32014114

Wael K. Hanna obtained his PhD from Information Systems Department, Faculty of Computer science and Information Systems, Mansoura University, Egypt. He obtained his B.S. in computer sciences and information systems from Sadat Academy in 2004 and he got M.SC degree in computer sciences and information systems from Sadat Academy for Management Sciences in 2011. Now he is a lecturer in Computer science and Information Systems Department at Sadat Academy for Management Sciences, Cairo, Egypt. His research interest is the Web Searching, Information Systems, Web mining, Medical Data Mining, Neutrosophic and Fuzzy Logic. He had published many articles. He is a reviewer for IGI Global and an editor for Inderscience.

Nouran M. Radwan obtained her PhD from Information Systems Department, Faculty of Computer science and Information Systems, Mansoura University, Egypt. She obtained his B.S. in computer sciences and information systems from Sadat Academy in 2004 and She got M.SC degree in computer sciences and information systems from Arab Academy for Science and Technology in 2011. Now she is a lecturer in Computer science and Information Systems Department at Sadat Academy for Management Sciences, Cairo, Egypt. Her research interest is Information Systems, Data Mining, Neutrosophic and Fuzzy Logic.