THE PREDICTION OF POSITIVE COVID-19 CASES IN MALAYSIA USING EXPONENTIAL SMOOTHING, PARABOLA AND CUBIC LEAST SQUARE METHOD

Nazri Asmawi Bin Jamaludin [1], Amiruddin Bin Ab Aziz[1], Nur Afriza Binti Baki[1], Bahari A.R.[2]
[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia
[2]School of Mechanical Engineering, Universiti Teknologi MARA, Malaysia

amiru2830@uitm.edu.my

ABSTRACT

Coronavirus Disease 2019 is the biggest problem that Malaysia needs to face. People who affected by coronavirus are the biggest challenge that we need to prevent and control. Therefore, it can be seen that people in Malaysia are mostly need to be more worry about COVID -19. However according to data provided. There has been a significant decreasing and also increasing in positive COVID-19 cases from January 2021 until October 2021. People cannot go to work or do their daily activities because of the pandemic COVID-19. This virus was spreading widely month by month also making people worry and carefully to prevent themselves from being affected. To overcome this problem, we need to predict the coming cases for us to be ready to live in this situation. In this article explains the best method to predict of COVID-19 in Malaysia by using the exponential smoothing method, parabola least square method and cubic least square method. The data of COVID-19 cases in Malaysia from January to October in a year 2021 is obtained from the Ministry of Health Malaysia and World Health Organization official website. The data from January to October in the year 2021 is calculated using all three methods, and the result is compared to the actual data. All methods' error is calculated using relative error. The method that produces the fewest errors is selected as the best.

Keywords: Covid 19, exponential smoothing method, parabola least square method, cubic least square method

1. INTRODUCTION

On 31 December 2019, the first case of Coronavirus 2019 (COVID-19) was reported at China in Wuhan. After that, virus was spreading widely around many countries turning it into pandemic. The number cases of COVID-19 grew fastly internationally a few months later. on 25 January 2020, Coronavirus disease incidence was discovered in Malaysia for the first time, among three Chinese citizens who entered the country across Singapore on 23 January 2020. The first outbreak commenced on the first incidence day until 16 February 2020. During that time, 22 confirmed cases were recorded with eight recovered patients and zero deaths, with only the most basic outbreak control procedures in place. However, on 27 February 2020, a new case involving two positive cases signaled the start of the second outbreak in Malaysia. During second outbreak, there was no significant increase in positive cases. Nevertheless, on 15 March 2020, the number of new cases jumped to 190 from 35 the day before. This alarming rate was caused by the discovery of huge groups of susceptible people who had come in contact with infected people. The dramatic increase in the outbreak's severity and rapid expansion changed the outbreak's Malaysian context from under control to dangerous. On March 18, 2020, the Movement Control Order (MCO) will go into action [1]. In such a difficult situation, identifying the COVID-19 infectious trend in Malaysia is critical for determining the pandemic's severity, as indicated by the massive increase in cases from three to 2766 in only 67 days. Estimating infectious counts over time allows researchers to have a better awareness of Malaysia's ongoing epidemiological status and gain insight into the measurable impact of outbreak control strategies. Therefore, in this study I am going to focus on the positive cases of Coronavirus disease 2019 in Malaysia.

The COVID-19 outbreak was studied from a Malaysian perspective, as well as mental health care during the outbreak. COVID-19 has no vaccination or specific treatment at that time. As a result, taking precautions is crucial if the SARS-CoV-2 infection is to be prevented and controlled. Preparedness should be a main focus for future pandemic outbreaks. [2]. Malaysia has the most instances among Southeast Asian countries. Despite the global spread of the COVID-19 pandemic, the number of patients in Malaysia is decreasing, with excellent recovery rates and low fatality rates. This viewpoint will address Malaysia's targeted containment measures, which have shown to be effective in controlling the spread of COVID-19 thus far [3].

In the absence of scientific and medical answers to COVID-19's many puzzles, In such critical circumstances, it is vital to acknowledge the need of well-implemented public health measures to avoid the virus's uncontrolled spread and multiplication into the human living environment, which would be impossible to combat. In order to combat the COVID-19 spread, Malaysia has chosen a unique focused approach [3].

Since exponential smoothing methods have historically been implemented without reference to a statistical framework, other methods to overcome the method selection problem have been created. One such method is prediction validation. The fitting sample and the validation sample are separated from the rest of the sample. The fitting sample is commonly used with a sum of squared one-step ahead prediction error criterion to discover correct value for the smoothing parameters [4].

The Single Exponential Smoothing method shows that as observed data get older, weighting drops exponentially. Value is given a higher weight later than the value of observation is longer. This method calculates an exponential weighted moving average of all previous observations' values. This strategy is unaffected by the season or the trend. The equation is as follows:

$$S_{t+1}=\alpha y_t+(1-\alpha)S_t \quad (1)$$

n the equation, to predict the value of the next period, the necessary demand data from previous periods and forecasting the previous period [5].

They propose a simple least squares approach to the problem of determining numerical values for parameters in differential equations so that the solution best fits some observable data in this paper. The method entails fitting the given data using least squares using cubic spline functions with knots selected interactively, and then determining the parameters using the least squares solution of the differential equation sampled at a number of points.

Calculate the "best straight line" using the least squares method. The most general solution, as well as the conditions under which previously determined specific solutions remain valid, are discovered. The Least-Squares Cubic solution is proven to yield the "ideal" slope. An example is provided to illustrate the procedure. The ideal slope is shown to be not necessarily bound by the values obtained from x on y and y on x regressions.

The least square methodology is a mathematical method for finding the "best fit" line to data that only takes some calculus and linear algebra. They also stated that identifying best fit would require a rapid study of probability and statistics. A closer examination of the facts reveals that is capable of establishing. They further claimed that it could determine the best fit provided by any finite linear combination of specified functions [6].

Increasing the degree of fit first appears to help, up to about degree 6, but then it becomes increasingly wavy. Although the residual is still reduced, it is evident that this is "fitting noise," and the inaccuracy actually worsens when compared to the "true" underlying cubic model. Overfitting is the term for this issue [7].

The relative error is defined as the ratio of a measurement's absolute error to the measurement being taken (RE). In other words, the type of error is determined by the size of the thing being measured. RE is expressed as a percentage and has no units [8]. While comparing errors, we use a reference model's forecast as a foundation, and when calculating errors, we use some forecast measure from a reference base model [9].

Relative Error is calculated as below:

$$Relative\ error, = \left|\frac{\sum X - \sum \bar{X}}{\sum X}\right| \tag{2}$$

The methods of prediction in this research which are exponential smoothing method (ESM), parabola least square method (PLSM) and cubic least square method (CLSM) were compared using relative error. The solution with least error is the best method to be used in predicts the cases of COVID-19. The method will benefit the government to predict the positive cases in order to beware and control of coronavirus spread thus preventing coronavirus from becoming uncontrollable.

2. RESEARCH METHODOLOGY

The research steps are clearly described one by one in a form of one step each.

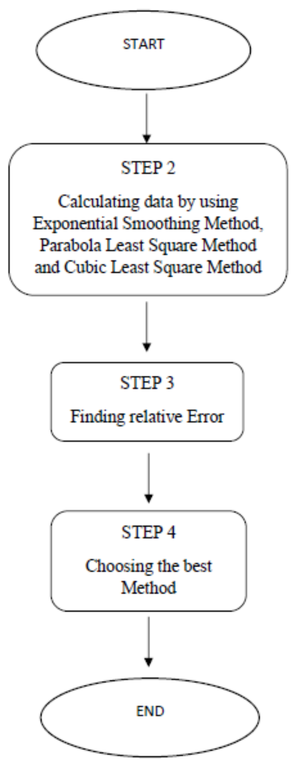

Figure 1: Flowchart of the Research Method

The type of data required for the study is monthly secondary data in year 2021. For a precise outcome, proper data is required. Data will be taken directly from a reliable and trusted source to ensure reliability and accuracy which is the World Health Organization official portal. This data will then be compared to data obtained from another source to ensure that the data used is highly accurate.

Secondly**,** the collected data will be applied into the three methods which are the exponential smoothing method, parabola least square method and cubic least square method to get the result. Based on the study it is found that the exponential smoothing method is a straight forward but reliable way to predicting. A frequent way for establishing the method suited for a specific time series is prediction validation on a withheld portion of the sample using criteria such as the mean absolute percentage error. From this, there are similarity that we also can apply for the data to predict the data cases. Microsoft excel will be used for the calculation.

Next, from the solutions need to find the relative error is estimated using the following formula for ESM, PLSM, and CLSM:

$$Relative\ error, \varepsilon = \left|\frac{\sum x - \sum \bar{x}}{\sum x}\right|$$

Where

$$x = actual\ value$$

$$\bar{x} = estimated\ value$$

This error will be calculated by using Microsoft Excel.

Finally, the relative errors from the three methods above have been plotted to get an accurate figure compared to the actual value.

3. RESULTS AND DISCUSSION

Table 1: The results of the Exponential Smoothing Method

X(Month)	Actual value	ESM	PLSM	CLSM
1	101949	101949	12310.79	163112.09
2	85793	89024.2	66248.99	15941.91
3	44748	53603.24	118548.23	-7174.49
4	63213	61291.05	169208.51	57850.67
5	163644	143173.4	218229.83	175105.17
6	179622	172332.3	265612.19	308676.79
7	361293	323500.9	311355.59	422653.31
8	632982	571085.8	355460.03	481122.51
9	499441	513770	397925.51	448172.17
10	220968	279528.4	438752.03	287890.07

The Table 1 above displayed the actual value which is from January to October 2021 and the forecasting value or expected value calculated from the ESM, PLSM and CLSM. Then, from the results above, the relative errors from the three methods have been plotted to figure out which method calculated have least errors compared to the actual value.

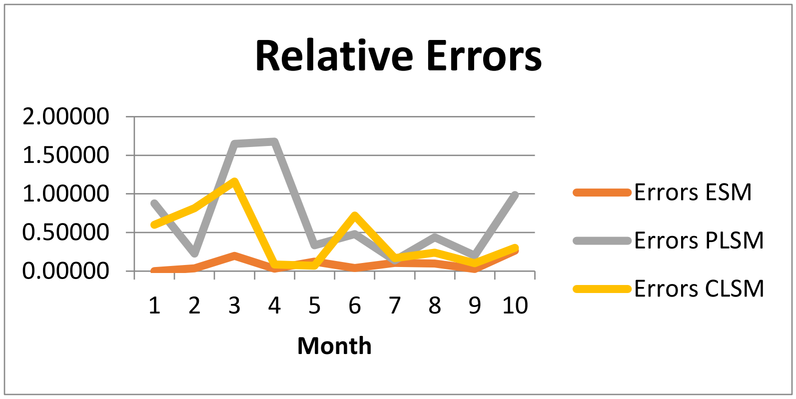

Figure 2: Graph of Error Analysis

From the Figure 2 above shows that the errors of expected value from exponential smoothing method have the least errors and the most accurate compared to other two methods. Meanwhile the parabola least square method have the largest errors and less accurate.

The total sum of relative errors for every method will be determined to do the error analysis. The table displayed the total summation of relative error obtained by each method.

	Exponential Smoothing	Parabola Least Square	Cubic Least Square
Total Sum of Relative Error	0.92773	7.01089	4.26306

Table 2: Total summation of Relative Errors

All three methods can be used to forecast COVID-19 positive cases in Malaysia based on the results obtained. However, by interpolating all three results from each method and also the actual value in the same graph, the value obtained from parabola Least Square method is the furthest from the actual value. The parabola least square method has the largest relative error and less accuracy with the other methods that used in this study after calculating based on calculating the relative error for all three methods. On the other hand, the expected value from exponential smoothing method is very close to the actual value. Therefore, the best method for prediction for cases of COVID-19 in Malaysia is determined by calculation the relative error. The exponential smoothing method has a smaller relative error which is 0.92773. It is proven that exponential smoothing method is the best method to be used to predict the number cases of COVID-19 in Malaysia.

4. CONCLUSION AND RECOMMENDATION

From the error analysis, data obtained from the exponential smoothing method has the smallest error which is 0.92773 and from the graph plotted the data is closer to the data compared to data obtained by using Parabola and Cubic Least Square Method.

According to the data analysis for all three methods, the parabola least square method has the biggest relative error which is the relative error is 7.01089, making it the least relevant method compare to other methods that used to calculate the relative error for predicting the number of positive COVID-19 cases in Malaysia. The data acquired using the Parabola Least Square Method is the furthest from the original data, as shown in the graph.

In conclusion, the exponential smoothing method (ESM), when compared to the parabola and cubic least method, is the best method for estimating the number of COVID-19 positive cases in Malaysia in the future since it has the smallest relative error and the data collected is the closest to the actual data.

REFRERENCES

[1] Ariffin, M. R. K., Gopal, K., Krishnarajah, I., Che Ilias, I. S., Adam, M. B., Arasan, J., … Mohammad Sham, N. (2021). Mathematical epidemiologic and simulation modelling of first wave COVID-19 in Malaysia. Scientific Reports..

[2] Elengoe, A. (2020). COVID-19 outbreak in Malaysia. *Osong public health and research perspectives*, *11*(3), 93. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7258884/.

[3] Ganasegeran, K., Ch'ng, A. S. H., & Looi, I. (2020). COVID-19 in Malaysia: Crucial measures in critical times. Journal of Global Health, 10(2). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7561271/.

[4] Qiao, L., Liu, D., Yuan, X., Wang, Q., & Ma, Q. (2020). Generation and prediction of construction and demolitionwaste using exponential smoothing method: A case study of Shandong Province, China. Sustainability (Switzerland). https://doi.org/10.3390/su12125094.

[5] Rachmat, R., & Suhartono, S. (2020). Comparative Analysis of Single Exponential Smoothing and Holt's Method for Quality of Hospital Services Forecasting in General Hospital. Bulletin of Computer Science and Electrical Engineering, 1(2), 80-86. http://bcsee.org/index.php/bcsee/article/view/comparative-analysis-of-single-exponential-smoothing-and-holts-m.

[6] Curry, M. (2014). Partial Least Squares Methodology for Analysis : a Primer. University of Manchester.

[7] Lei, I. L., Teh, P. L., & Si, Y. W. (2021). Direct least squares fitting of ellipses segmentation and prioritized rules classification for curve-shaped chart patterns. Applied Soft Computing. https://doi.org/10.1016/j.asoc.2021.107363.

[8] Chen, K., Lin, Y., Wang, Z., & Ying, Z. (2016). Least product relative error estimation. Journal of Multivariate Analysis. https://doi.org/10.1016/j.jmva.2015.10.017.

[9] Feroze, N. (2020). Forecasting the patterns of COVID-19 and causal impacts of lockdown in top five affected countries using Bayesian Structural Time Series Models. Chaos, Solitons and Fractals. https://doi.org/10.1016/j.chaos.2020.110196.