2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)
April 09-11. 2021, ISTTS Surabaya, Indonesia

Time Series Forecasting for the Spread of Covid-19 in Indonesia Using Curve Fitting

2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT) | 978-1-6654-0514-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/EIConCIT50028.2021.9431936

1st Arif Kurniawan
Magister of Informatic Engineering
University of Islamic State Maulana Malik Ibrahim
State Vocational High School 9
Malang, Indonesia
19841002@student.uin-malang.ac.id

2nd Fachrul Kurniawan
Magister of Informatic Engineering
University of Islamic State Maulana Malik Ibrahim
Malang, Indonesia
fachrulk@ti.uin-malang.ac.id

***Abstract*—The spread of Covid-19 has a huge impact on human life around the world, especially in Indonesia. Until now, the estimation of when this pandemic will end is still a big question. Therefore, a prediction is needed as the basis for the policies of the interested parties. In this paper, the Curve Fitting method is used to estimate the period and total cases of the spread of Covid-19 based on time series data for positive case growth without considering the rate of recovery, lockdown policy, and so on. With the proposed method and calculated based on the normal curve, the estimated end of a pandemic is at day seven hundred and fifty or about two years since the start of the pandemic. So that this estimate can be used as a basis by the government to make policies, for example when schools can be conducted face-to-face or other policies related to community needs.**

Keywords—covid-19, curve fitting, prediction

I. Introduction

The increasing number of health crises that have an impact on the world economy has practically forced all countries in the world to retreat with predetermined strategic plans which are then replaced by emergency response policies by mobilizing all resources to cope with the Covid-19 pandemic [1], [2]. Many questions arise that must be answered to anticipate or consider the policies of interested parties. For example, about when the covid-19 pandemic will end, or at least to find out when the pandemic shows a downward trend, with the hope that can give suggestions and recommendations for the government for making policies in not only economic but also other various fields, such as education [3], [4].

Prediction or forecasting is the art and science of predicting future events by taking historical data and projecting it into the future using mathematical models [5]. Predictions will certainly not be possible to draw true conclusions, but using a mathematical model and supported by good historical data will result in a high degree of precision [6].

Therefore, this paper will discuss how to mathematically forecast when the Covid-19 pandemic will end. To do this, the curve fitting method will be used, which uses two functions, namely sigmoid and Gaussian to fit the growth data of the spread of Covid-19. A programming language with complete library support will help in mathematical calculations and can be used to visualize data and information. With mathematical functions implemented in programming, it is hoped that it can display a visualization that shows what day the pandemic will end.

II. Data Set

The time-series dataset of Covid-19 was obtained from the Covid-19 Indonesian Research Repository website. The website recapitulates data from the *Kawal Covid-19* publications portal, where this portal is an information hub, that updates relevant information regarding the development of Covid-19 in Indonesia on daily basis [7]. We try to take updated data from March 2, 2020, until December 19, 2020. With time-series data for approximately ten months, it is expected that it will be sufficient to make forecasts that are closer to real conditions [8]. Fig. 1 and 2 represent a data set of the growth in positive cases of Covid-19. Fig. 1 is the data head, showing the onset of the pandemic, and Fig. 2 is the data tail, which shows the most recent data.

III. Method

Curve fitting, moreover known as regression analysis, is utilized to discover the "finest fit" line or curve for an arrangement of data points [9], [10]. Most of the time, the curve fit will deliver a condition that can be utilized to discover points anywhere along the curve. In a few cases, you will not be concerned about finding a condition. Instep, you will fair want to utilize a curve fit to smooth the data and move forward the appearance of the plot [11]. Several curves can be used in curve fitting, it's just that in this experiment only two curves will be used, namely the Sigmoid curve and the Gaussian curve.

	date	totalconfirmed	dailyconfirmed
0	3/2/2020	2	2
1	3/3/2020	2	0
2	3/4/2020	2	0
3	3/5/2020	2	0
4	3/6/2020	4	2

Fig. 1. Screenshot of the data head

	date	totalconfirmed	dailyconfirmed
288	12/15/2020	629429	6120
289	12/16/2020	636154	6725
290	12/17/2020	643508	7354
291	12/18/2020	650197	6689
292	12/19/2020	657948	7751

Fig. 2. Screenshot of data tail

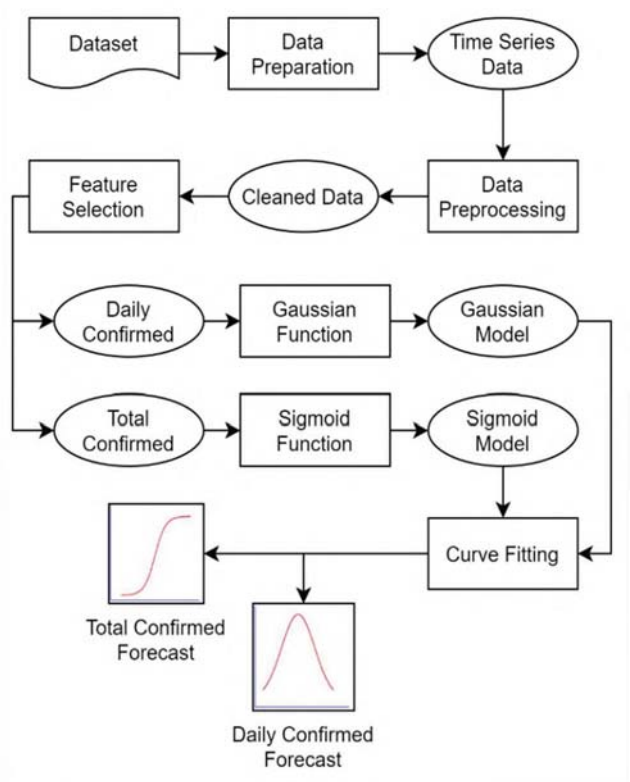

Fig. 3. Block Diagram of Proposed Method

Fig. 3 shows the proposed method. After doing data preparation and data preprocessing, we take only two features. The dataset used is data in a time series format, where in addition to the date attribute there are also two other attributes which are features used to make predictions. The first feature is cumulative data on positive Covid-19 patients and the second feature is data on the increase in cases per day. Each feature is entered into a function so that each function produces a model. Then the resulting model is used as the main parameter in the curve fitting function and produces the forecasted curve.

A. Sigmoid Curve

A sigmoid function is a mathematical function that generates a curve having an "S" shape [12]. The sigmoid curve is the same as the Growth curve. Growth curves are utilized to demonstrate the growth of an individual/organism or populace (plants or creatures) over time. Growth estimations (subordinate variable) can be body mass, body length, or mass/length of distinctive parts of the body. Growth curves can have different growth designs. On the off chance that there's an upper restrain, for illustration, since of natural assets, growth continues sigmoidally and asymptotically. Within the starting stage, growth is relatively slow but quickens from there on and diminishes within the last stage, drawing nearer the upper asymptote [13].

Various things utilized the sigmoid curve in growth analysis, and numerous growth models have been proposed. Among the foremost habitually utilized are logistic, *Richards and Gompertz* models [13]. So, this paper will use logistics to draw the growth of the Covid-19 case. And the Sigmoid curve equation is as follows:

$$f_{a,b,c,d}(x) = \frac{a}{1+be^{-cx+d}} \quad (1)$$

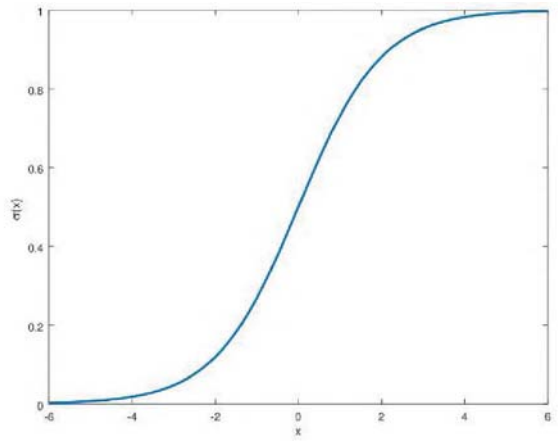

Fig. 4. Logistic Sigmoid Curve

The parameter a is the curve's maximum value or the max number of a cumulative confirmed case, b is the model constant, c is the steepness of the curve or the growth rate factor, and d is the day of the peak of the pandemic on an arbitrary timescale (i.e. the sigmoid's midpoint) [14]. Fig. 4 shows a graph of the Sigmoid function starting with a gently sloping line, then ascending and ending with a gentle slope.

B. Gaussian Curve

The Gaussian curve is the same as a normal distribution. The normal distribution is a theoretical symmetrical conveyance utilized to form comparisons among scores or to create other sorts of factual choices. The shape of this conveyance is regularly alluded to as "bell-shaped" or colloquially called the chime curve"[15]. This shape infers that the larger part of scores lie near to the center of the dispersion, and as scores float from the center, their recurrence diminishes [16]. And the Gaussian curve equation is as follows:

$$g_{a,b,c}(x) = \frac{a}{c}e^{-\frac{(x-b)^2}{c^2}} \quad (2)$$

The parameter a is the height of the curve's peak or max daily confirmed, b is the position of the center of the peak or the day when daily cases were highest, and c (the standard deviation) controls the width of the "bell" [17]. Fig. 5. shows a graph of the Gaussian function and is a normal distribution that usually occurs.

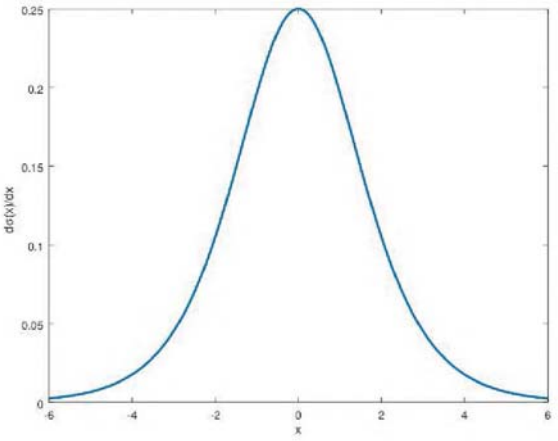

Fig. 5. Gaussian or Bell Curve

C. Data Visualization

Data visualization means presenting data in a graphical or pictorial form which makes the information easy to understand [18]. It helps to explain facts and determine courses of action. It will benefit any field of study that requires innovative ways of presenting large, complex information. The advent of computer graphics has shaped modern visualization.

IV. Discussion

From the available data set, which is in the form of time-series data from the development of the number of cases and daily cases, data visualization is generated presented in Fig. 6. In Fig. 6 the left shows the development of the number of cases (cumulative) from the start of positive patients until the last day of the trial. It can be seen that the trend is still increasing in cases. Whereas in the right picture shows the development of daily cases, where if we observe the trend is still dominant up, even though in detail there is a decrease in a certain time frame but after that, it shows an increase again.

A. Cumulative Total Cases

To predict cumulative total case must be used growth or sigmoid curve on the available time-series data [19]. Based on equation 1, in the case of Sigmoid its need to estimate parameters a, b, c, d by curve fitting. Fig. 7. shows the sigmoid function in computer programming, where this function has five parameters as input, the Sigmoid formula as a process, and returns the calculation results of the formula as a Sigmoid model.

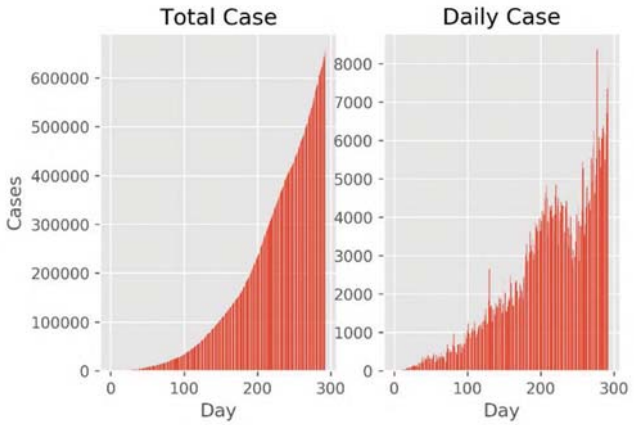

Fig. 6. Total case and the daily case of Covid-19 in Indonesia

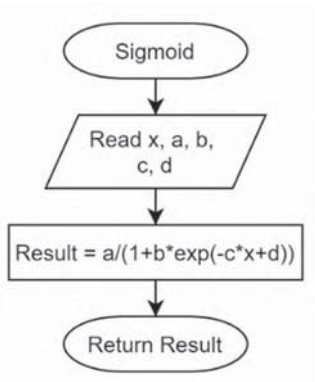

Fig. 7. Flowchart of Sigmoid function

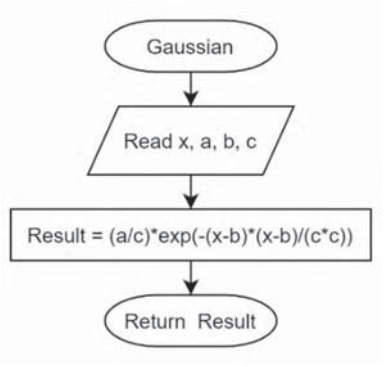

Fig. 8. Flowchart of Gaussian function

B. Daily Cases

To predict daily cases must be used Gaussian curve on the available time-series data [20]. Based on equation 2, in the case of Gaussian its need to estimate parameters a, b, c by curve fitting. Fig. 8. shows the Gaussian function in computer programming, where this function has four parameters as input, the Gaussian formula is a process and returns the calculation results of the formula as a Gaussian model.

V. Results

After revoking the sigmoid function and plotting the result, the predicted results of total cases can be seen in Fig. 9 that the prediction results show a curve that starts to slope on day 750 from the beginning of the case for positive patients. This shows that on that date there were no more new cases, therefore the curve tended to slope after an increase in numbers. Whereas in Fig. 10 is the result of daily case predictions, where the most daily cases or the peak occur around day 375 and after that, there is a decrease in the number of cases until there are no more cases on day 750. The prediction results show that the Mean Square Error (MSE) tends to be high. This is because the daily virus growth data is unstable, which is probably because the data recorded by the government does not represent the real condition [21]. Under normal conditions, biological growth usually occurs regularly [22].

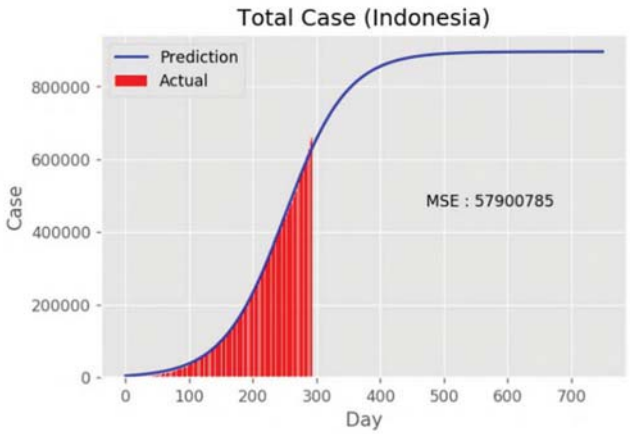

Fig. 9. Prediction result of cumulative total cases

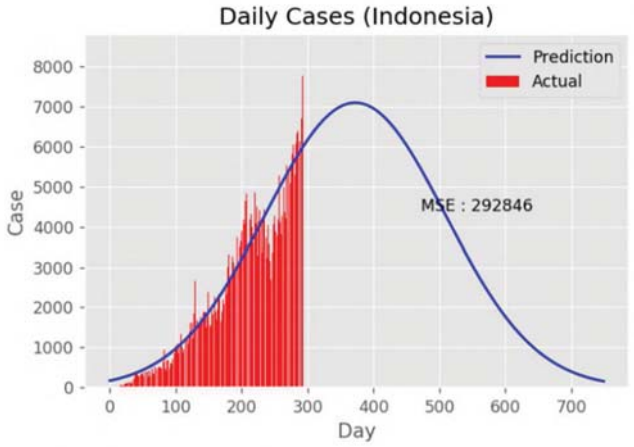

Fig. 10. Prediction result of daily cases

CONCLUSION

By using the curve fitting method, it turns out to be able to predict the development of the case of the spread of Covid-19. Even so, the predictions made are still limited to utilizing data on positive case growth and do not involve other variables such as the number of patients who have died, the number of patients who have recovered, and possibly also government policies in handling pandemics such as lockdown policy, use of vaccines and so on. From the prediction, we can conclude that the spread of Covid-19 is still ongoing today. Even today it has not yet reached its peak and the final prediction of a pandemic is far from now. This assumption is important to become the basis for the government to make policies in the fields of health, economy, education, social, and so on.

REFERENCES

[1] A. Pradesha, S. Amaliah, A. Noegroho, and J. Thurlow, The Cost Of COVID-19 on the Indonesian Economy: A Social Accounting Matrix (SAM) Multiplier Approach, Intl Food., no. June. 2020.

[2] R. Djalante et al., "Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020," Prog. Disaster Sci., vol. 6, no. April, p. 100091, 2020, doi: 10.1016/j.pdisas.-2020.100091.

[3] M. Sukmana, M. Aminuddin, and D. Nopriyanto, "Indonesian Government Response In COVID-19 Disaster Prevention," East Afrian Sch. J. Med. Sci., vol. 3, no. 3, pp. 81–6, 2020, doi: 10.36349/EASMS.-2020.v03i03.025.

[4] F. Kurniawan, F. A. Dwiyanto, G. Budiprasetyo, and M. I. Esha, "Sustainable Development for Smart City Education Based on Features and Parameters Assessment Modelling Slope One Algorithm," J. Green Eng., vol. 10, no. 10, pp. 9801–9811, 2020.

[5] S. Makridakis, R. J. Hyndman, and F. Petropoulos, "Forecasting in Social Settings: The State of The Art," Int. J. Forecast., vol. 36, no. 1, pp. 15–28, 2020, doi: 10.1016/j.ijforecast.2019.05.011.

[6] L. Jia, K. Li, Y. Jiang, and X. Guo, "Prediction and Analysis of Coronavirus Disease 2019," arXiv, no. December. p. 2003.05447, 2020.

[7] I. Nadzir, "Data Activism and Covid-19 in Indonesia (Aktivisme Data Dan Covid-19 Di Indonesia)," J. Kependud. Indones., vol. 2902, no. Edisi Khusus Demografi dan COVID-19, pp. 105–108, 2020.

[8] L. Li et al., "Propagation Analysis and Prediction of the COVID-19," Infect. Dis. Model., vol. 5, pp. 282–292, 2020, doi: 10.1016/j.idm.-2020.03.002.

[9] M. Kowsher, M. Uddin, M. M. Moheuddin, and M. Y. Turaba, "Two New Regression and Curve Fitting Techniques Using Numerical Methods," Algorithms for Intelligent Systems, no. January. Springer., 2020, doi: 10.2139/ssrn.3590089.

[10] D. Grover, "Analysis of Various Exponential Curve Fitting: Linearization of Non-Linear Laws," Eur. J. Mol. Clin. Med., vol. 7, no. 7, pp. 3881–3889, 2020.

[11] H. Yonar, A. Yonar, M. A. Tekindal, and M. Tekindal, "Modeling And Forecasting for The Number of Cases of the COVID-19 Pandemic with The Curve Estimation Models, The Box-Jenkins And Exponential Smoothing Methods," Eurasian J. Med. Oncol., vol. 4, no. 2, pp. 160–165, 2020, doi: 10.14744/ejmo.2020.28273.

[12] F. Salim and N. A. Abu, "An s-curve model on the maximum predictive pricing of used cars," Eur. J. Mol. Clin. Med., vol. 7, no. 3, pp. 907–921, 2020.

[13] J. Ukalska and S. Jastrzębowski, "Sigmoid Growth Curves, A New Approach to Study the Dynamics of The Epicotyl Emergence of Oak," Folia For. Pol., vol. 61, no. 1, pp. 30–41, 2019, doi: 10.2478/ffp-2019-0003.

[14] L. Cao, P. J. Shi, L. Li, and G. Chen, "A New Flexible Sigmoidal Growth Model," Symmetry (Basel)., vol. 11, no. 2, pp. 1–20, 2019, doi: 10.3390/sym11020204.

[15] J. Schüttler, R. Schlickeiser, F. Schlickeiser, and M. Kröger, "Covid-19 Predictions Using A Gauss Model, Based on Data from April 2," Physics (College. Park. Md)., vol. 2, no. 2, pp. 197–212, 2020, doi: 10.1101/2020.04.06.20055830.

[16] N. B. Salah, "Some Properties Of Q-Gaussian Distributions," arXiv preprint. pp. 1–17, 2021.

[17] Y. Nishimoto and K. Inoue, "Curve-Fitting Approach For COVID-19 Data and Its Physical Background," medRxiv. 2020.

[18] A. Unwin, "Why is Data Visualization Important? What is Important in Data Visualization?," Harvard Data Sci. Rev., no. 2, pp. 1–7, 2020, doi: 10.1162/99608f92.8ae4d525.

[19] P. Chilinski and R. Silva, "Neural Likelihoods Via Cumulative Distribution Functions," in Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence, UAI, 2020, pp. 420–429.

[20] D. Toth and B. Jones, "Against the Norm: Modelling Daily Stock Returns with The Laplace Distribution," arXiv. pp. 1–5, 2019.

[21] C. Gao, D. Shi, and A. Maydeu-Olivares, "Estimating the Maximum Likelihood Root Mean Square Error of Approximation (RMSEA) with Non-normal Data: A Monte-Carlo Study," Struct. Equ. Model., vol. 27, no. 2, pp. 192–201, 2020, doi: 10.1080/10705511.2019.1637741.

[22] A. Millet, "A Universal Model for The Log-Normal Distribution of Elasticity in Polymeric Gels and Its Relevance to Mechanical Signature of Biological Tissues," Biology (Basel)., vol. 10, no. 1, pp. 1–9, 2021, doi: 10.3390/biology10010064.