A Comparison of Newton's Divided Differences Interpolation and a Cubic Spline in Predicting the Poverty Rate of West Java

Sri Purwani, Rizki Apriva Hidayana, Viona Prisyella Balqis, and Sukono

Abstract—Poverty is the condition of having few material possession or little income. It can have various economic, social, and political causes and effects and has served as a major impediment to Indonesia's national development process. Using data on the rising poverty rate in West Java up to March 2021, we can use interpolation to construct a function that describes the pattern of the data. The function can then be used to predict poverty rate data beyond the scope of the already established data. There are several formulas of interpolation whose graphs pass through the interpolation points, such as Lagrange, Newton's divided differences and cubic splines. We aim to compare Newton's divided differences interpolation and a cubic spline in the context of predicting West Java poverty rates. The results obtained show that the interpolation using a cubic spline is superior to Newton's interpolation polynomial because a cubic spline results in smaller errors.

Index Terms: Poverty rate, interpolation, cubic spline, Newton's interpolation polynomial

I. INTRODUCTION

THE poverty rate is a risk factor for the overall well-being of a country and is reflected by the pattern of consumption (which relates to material, such as food and non-food) and income rate [1]. According to Statistics Indonesia, the poverty rate in West Java has been increasing as of March 2021.

Poverty rate data can be analysed using computational methods, one of which is prediction. Prediction, which is the process of estimating unknown data based on the available or historical data, can be performed by using interpolation.

Manuscript received 24 May 2022; revised 20 August 2023.

Sri Purwani is a lecturer in the Mathematics Department, Universitas Padjadjaran, Jl. Raya Bandung-Sumedang km 21 Jatinangor, Sumedang 45363. Indonesia (corresponding author to provide e-mail: sri.purwani@unpad.ac.id).

Rizki Apriva Hidayana is a graduate of Magister of Mathematics, Universitas Padjadjaran, Jl. Raya Bandung-Sumedang km 21 Jatinangor, Sumedang 45363. Indonesia (e-mail: rizki20011@mail.unpad.ac.id).

Viona Prisyella Balqis is a graduate of Magister of Mathematics, Universitas Padjadjaran, Jl. Raya Bandung-Sumedang km 21 Jatinangor, Sumedang 45363. Indonesia (e-mail: viona20004@mail.unpad.ac.id).

Sukono is a Profesor in the Mathematics Department, Universitas Padjadjaran, Jl. Raya Bandung-Sumedang km 21 Jatinangor, Sumedang 45363. Indonesia (e-mail: sukono@unpad.ac.id).

Interpolation is a way of getting or calculating a function whose graph passes through a group of points provided. It can also be defined as a method of 'guessing' the value of data based on other information [2]. It generally uses the interpolating function from a restricted class of functions, the most commonly used class of which is polynomials. It has also been used to optimise the nonparametric estimation and prediction of loss square problems [3]. In addition, it includes formulas such as Lagrange, Newton's divided differences, and a cubic spline, all of which generate a smooth function [4]. A cubic spline is a piecewise polynomial function. It uses a spline function for the interpolating function. It is hence continuous over each subinterval. Moreover, as in the construction, we impose the spline and its first and second derivatives to be continuous at the same point between two adjacent subintervals, meaning it is then continuous throughout the interval [5].

James et al. proved that cubic spline interpolation was far more effective than linear interpolation in some cases [6]. Behforooz et al. found that the final condition of the cubic spline function was derived using the Newtonian integral method in solving the problem of linear complementarity [7]. Asen et al. proved local square convergence by seeing it as Newton's method [8]. Linear complementarity problems were solved by using a kind of Newton's method [9]. Hoschek and Schwanecke discussed the linear method for interpolation [10]. Goonatilake and Ruiz discussed the comparison between polynomial interpolation using numerical versus statistical techniques [11]. Gasca and Sauer presented multivariate polynomials with Newton's approach [12], Syam presented higher-level predictor methods with cubic splines [13].

Based on the previous research, this study will assess the effectiveness of interpolation by using Newton's interpolation polynomial and a cubic spline on poverty rate data in West Java. In doing so, we aim to devise a better method for estimating and evaluating the function that represents the pattern of the data. By having the function, we can predict data other than the available ones, within or beyond the range of the data.

II. MATERIAL AND METHODS

The data used in this study refers to the poverty rate in West Java from 2010 to 2020. It is obtained from Statistics Indonesia on the website https://jabar.bps.go.id and was posted on November 6, 2021. The data is shown in Table I

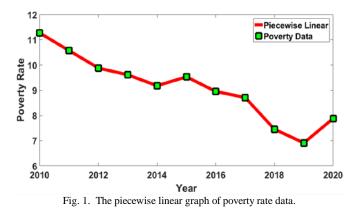
This work was supported in part by Universitas Padjadjaran through the Universitas Padjadjaran Internal Research Grant, Fiscal Year 2023, in the "Academic Leadership Grant (ALG)" program under Prof. Dr. Sukono, MM., M.Si., with contract number 1549/UN6.3.1/PT.00/2023.

	TABLE I	
POVERTY RATE IN WEST JAVA		
Year	Poverty Rate (percentage)	
2010	11.27	
2011	10.57	
2012	9.88	
2013	9.61	
2014	9.18	
2015	9.53	
2016	8.95	
2017	8.71	
2018	7.45	
2019	6.91	
2020	7.88	

as follows,

the following.

Piecewise linear interpolation is then applied to the data in order to connect each successive piece of data listed in Table I. This is shown in Fig. 1.



As interpolation, the graph passes through each data point; however, it is not smooth. Most data represent a smooth graph. We aim to construct a smooth curved graph interpolating the given data points while preserving the shape

of the piecewise linear interpolation. This will be discussed in

A. Newton's Divided Differences Interpolation Polynomial

Newton's divided differences interpolation polynomial, or Newton's interpolation polynomial for short, can be thought of as a rearrangement of the Lagrange polynomial for computational efficiency. This is because the former can reuse its lower-degree polynomial to construct a higherdegree polynomial, whereas the latter has to start from the beginning to construct a higher-degree polynomial. However, the latter is used intensively to understand and develop a theory.

Having n+1 data points denoted by (x_i, y_i) with i = 0, 1, 2, ..., n and $x_0, x_1, ..., x_n$ be distinct numbers, we can construct Newton's interpolation polynomial of degree less or equal to *n*. This polynomial can take the following form:

$$P_n(x) = f(x_0) + f[x_0, x_1](x - x_0) + \dots + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1)\dots(x - x_{n-1})$$
(1)

where $f[x_0, x_1, ..., x_i]$ are a form of divided differences defined as follows:

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}, f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}$$
$$f[x_0, x_1, \dots, x_n] = \frac{f[x_1, x_2, \dots, x_n] - f[x_0, x_1, \dots, x_{n-1}]}{x_n - x_0}$$

The relation between the lower-degree Newton interpolation polynomial and the higher-degree one can then be stated as follows:

$$P_n(x) = P_{n-1}(x) + f[x_0, x_1, ..., x_n](x - x_0)(x - x_1)...(x - x_{n-1})$$
(2)

Equations (1) and (2) can be efficiently performed by using nested multiplication, which will be shown later in (9). Furthermore, Equation (2) allows us to construct a higherdegree Newton's interpolation polynomial by only adding the last term to the lower-degree one. This significantly saves on computation time.

This method produces a smooth function over the entire interval. However, for a polynomial of degree more than 5, it tends to be highly oscillated or changes significantly from the two successive data values near the ends of the interval (see Fig. 2). This unavoidably results in large errors of the function values within these parts of the interval. It is this issue that motivates the following method of interpolation.

B. Cubic Spline Interpolation

We aim to have interpolation that is smooth but which maintains its shape for the given data. This leads to the use of spline interpolation.

For the given **n** data points,
$$\{(x_i, y_i)\}_{i=1}^n$$
, we assume that

$$a = x_1 < x_2 < \dots < x_n = b \tag{3}$$

and look for a function $\xi(x)$ defined on the interval [a,b] interpolating the data points

$$\xi(x_i) = y_i, \quad i = 1, 2, ..., n$$
 (4)

In order to make $\xi(x)$ smooth, we require $\xi'(x)$ and $\xi''(x)$ to be continuous. We then aim to retain the common shape of the piecewise linear function relating the data points (x_i, y_i) (see Fig. 1). This is done by asking the derivative $\xi'(x)$ not to change too rapidly between node points, which means we require the second derivative $\xi''(x)$ to be as small as possible. There is a unique solution $\xi(x)$ to this problem called the natural cubic spline that interpolates the data $\{(x_i, y_i)\}_{i=1}^n$ and satisfies the following:

F1. $\xi(x)$ is a polynomial of degree ≤ 3 on each subinterval $[x_{k-1}, x_k]$, k = 2, 3, ..., n;

F2. $\xi(x), \xi'(x), \xi''(x)$ are continuous for $a \le x \le b$; F3. $\xi''(x_1) = \xi''(x_n) = 0$.

To construct the natural cubic spline, we introduce the variables $S_1, S_2, ..., S_n$ with

$$S_i = \xi''(x_i), \quad i = 1, 2, ..., n.$$

and state the spline $\xi(x)$ in terms of S_i . We then construct a system of linear equations from which the values S_i can be found.

We ensure that $\xi(x)$ is cubic on each interval $[x_{k-1}, x_k]$ so that its second derivative $\xi''(x)$ is linear on the interval. Constructing a linear function requires the following values at two points:

$$\xi''(x_{k-1}) = S_{k-1}, \quad \xi''(x_k) = S_k,$$

which gives us

$$\xi''(x) = \frac{(x_k - x)S_{k-1} + (x - x_{k-1})S_k}{x_k - x_{k-1}}, \ x_{k-1} \le x \le x_k$$
(5)

The next step is to construct the second antiderivative of $\xi''(x)$ on $[x_{k-1}, x_k]$, for which we use the following interpolation properties:

 $\xi(x_{k-1}) = y_{k-1}, \ \xi(x_k) = y_k$ This gives the cubic polynomial [14]

$$\xi(x) = \frac{(x_{k} - x)^{3} S_{k-1} + (x - x_{k-1})^{3} S_{k}}{6(x_{k} - x_{k-1})} + \frac{(x_{k} - x) y_{k-1} + (x - x_{k-1}) y_{k}}{x_{k} - x_{k-1}} - \frac{1}{6} (x_{k} - x_{k-1}) \left[(x_{k} - x) S_{k-1} + (x - x_{k-1}) S_{k} \right]$$
(6)

for $x_{k-1} \le x \le x_k$. Formula (6) applies to each of the intervals $[x_{k-1}, x_k]$, k = 2, 3, ..., n. This satisfies F1–F3 for the continuity of $\xi(x), \xi'(x), \xi''(x)$ on [a,b] and uses the following system of linear equations [14]:

$$\frac{1}{6} \left(\left(x_{k} - x_{k-1} \right) S_{k-1} + 2 \left(x_{k+1} - x_{k-1} \right) S_{k} + \left(x_{k+1} - x_{k} \right) S_{k+1} \right)$$
$$= \frac{y_{k+1} - y_{k}}{x_{k+1} - x_{k}} - \frac{y_{k} - y_{k-1}}{x_{k} - x_{k-1}}, \quad k = 2, 3, \dots, n-1.$$
(7)

This gives n-2 equations which together with the previous assumption (F3), $S_1 = S_n = 0$, produce the values $S_1, S_2, ..., S_n$ and the interpolating function $\xi(x)$.

The interpolating spline $\xi(x)$ is highly effective when we only have data points and only aim to have a seemingly smooth plot. However, for the spline interpolating a known function, we can then observe its accuracy.

Suppose f(x) is defined on the interval [a,b] and

consider the interpolation of f(x) at evenly spaced values of x. Let n > 1, which gives us the following:

$$x_k = a + (k-1)h, \quad h = \frac{b-a}{n-1}, \quad k = 1, 2, ..., n$$

We denote $\xi_n(x)$ to be the natural cubic spline that interpolates f(x) at $\{x_i\}_{i=1}^n$ and satisfies

$$\max_{a \le x \le b} \left| f\left(x\right) - \xi_n\left(x\right) \right| \le ch^2 \tag{8}$$

where *c* depends on f''(a), f''(b) and $\max_{a \le x \le b} \left| f^{(4)}(x) \right|$. The estimate $\xi_n(x)$ that does not converge rapidly is due to f''(x) being commonly nonzero at x = a and *b*. However, we have defined $\xi_n''(a) = \xi_n''(b) = 0$ (see F3).

Improving on $\xi_n(x)$ requires other functions of cubic spline $\xi(x)$ interpolating f(x). With the definition of node-point (3), $\xi(x)$ is a cubic spline on [a,b] if it is cubic on each subinterval $[x_{k-1}, x_k]$, and all $\xi(x), \xi'(x)$ and $\xi''(x)$ are continuous on [a,b].

Supposing we choose $\xi(x)$ to satisfy the interpolating condition (4) and obtain the data by evaluating f(x), we get the following:

$$y_i = f(x_i), \quad i = 1, 2, ..., n$$

We then take boundary conditions for $\xi(x)$ to form a better estimate of f(x). The boundary conditions can either

or

 $\xi''(x_1) = f''(x_1), \qquad \xi''(x_n) = f''(x_n)$

 $\xi'(x_1) = f'(x_1), \qquad \xi'(x_n) = f'(x_n)$

Either of these boundary conditions, together with (6) and (7) results in a unique interpolating spline $\xi(x)$, dependent on the condition chosen. In both cases, we can have ch^4 on the right-hand side of (8), where c depends on $\max_{a \le x \le b} |f^{(4)}(x)|$.

If we have no derivatives of f(x), we can apply another boundary condition to preserve the error bound of (8) to be proportional to h^4 . In particular, we assume we have

 $x_1 < u_1 < x_2$, $x_{n-1} < u_2 < x_n$ and assume that $f(u_1)$ and $f(u_2)$ are known. We then apply $\xi(x)$ in (6) and

$$\xi(u_1) = f(u_1), \qquad \xi(u_2) = f(u_2).$$

This results in two new equations in (7). This form of cubic spline is commonly preferable to the natural cubic spline and satisfies the not-a-knot interpolation boundary conditions. Another boundary condition was used by rational cubic or quadratic spline to construct interpolation curves preserving positivity [15]. Spline is also applied on a flat surface, at points to be connected, which produces a smooth curve [16].

C. Extrapolation

Extrapolation is the process of estimating function values beyond the limits of the observed data. To do this, it uses known data to estimate other data outside the range of the known data. In this research, extrapolation was carried out by using Newton's divided differences and cubic spline functions, as discussed in the previous sections.

III. RESULT AND DISCUSSION

A. Newton's Divided Differences Interpolation

The Newton formula (1), when applied to poverty data in Table I, is rewritten in the form of nested multiplication, which is given as follows:

$$P_{10}(x) = 11.27 + (x - 2010)(-0.70 + (x - 2011)(0.005 + (x - 2012)(0.068333 + (x - 2013)(-0.041250 + (x - 2014)(0.020917 + (x - 2015)(-0.009278 + (x - 2016)(0.003270 + (x - 2017)(-0.000931 + (x - 2018)(0.093952 - 0.000046x)))))))))))$$

$$(9)$$

This formula (9) is then plotted to give the graph of Newton's interpolation polynomial as shown in Fig. 2.

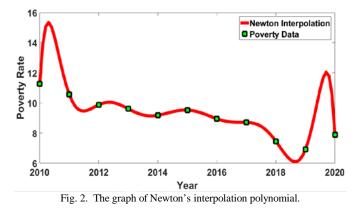


Fig. 2 shows that the graph of Newton interpolation exactly passes through the interpolation data. However, it greatly changes and results in large errors, especially on subintervals [2010,2011] and [2019,2020] [17]. Equation (9) is then used to predict the midpoints of poverty rate data which are given in Table II,

MIDDONIT DATA DOODUC	ED BY NEWTON INTERPOLATION
	Poverty Rate (percentage)
2010.5	14.39289102
2011.5	9.461331251
2012.5	10.01932900
2013.5	9.202499466
2014.5	9.415500870
2015.5	9.321175003
2016.5	8.731406638
2017.5	8.448451633
2018.5	6.212983735
2019.5	10.93261153

To compare the results we augment the midpoint data and the piecewise linear (Fig. 1) to the graph in Fig. 2 and obtain the graphs as shown in Fig. 3,

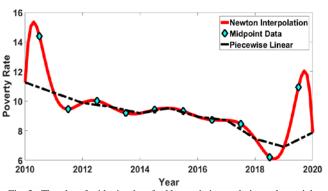
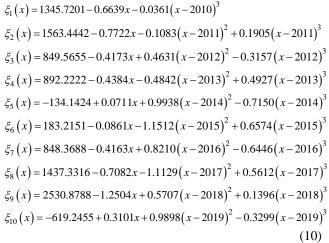


Fig. 3. The plot of midpoint data for Newton's interpolation polynomial.

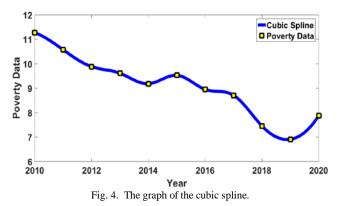
Fig. 3 shows that besides on subintervals [2010,2011] and [2019,2020], greater errors of midpoint data also occur on subintervals [2011,2012] and [2018,2019]. This makes the prediction of data points on those subintervals less accurate. To ensure more accurate predictions, then, other than the available ones, we need interpolation that fits smoothly to the data and has smaller errors through the interval. This will be discussed in the following section.

B. Cubic Spline Interpolation

The cubic spline is arranged in such a way as to have properties that allow for good prediction. Performing (6) and (7) on poverty data in Table I, gives us the following cubic spline:



Equation (10) is then plotted to form the graph of the cubic



spline as shown in Fig. 4.

Fig. 4 shows that the cubic spline produces a smooth function without significant errors through the entire interval [2010, 2020]. This agrees with the aim of its construction (see Section II Subsection B), which is to be smooth while preserving the pattern of the data. Therefore, we can expect to form more accurate predictions regarding poverty data by using a cubic spline. Using (10) then results in the midpoints of poverty rate data in West Java (see Table III).

Midpoint D	TABLE III ATA PRODUCED BY A CUBIC SPLINE
Year	Poverty Rate (percentage)
2010.5	10.9335
2011.5	10.1806
2012.5	9.7476
2013.5	9.3312
2014.5	9.3746
2015.5	9.2812
2016.5	8.8664
2017.5	8.1477
2018.5	6.9849
2019.5	7.2712

The midpoint data together with the piecewise linear (see Fig. 1) are then augmented to the previous plot (Fig. 4) to create the graph presented in Fig. 5.

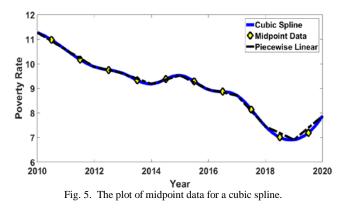
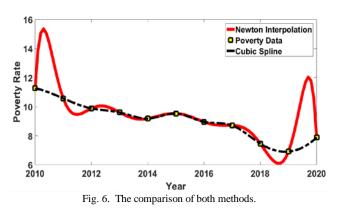


Fig. 5 shows that the cubic spline (10) and its midpoint data follow the pattern of poverty rate data. The errors of the cubic spline remain small through the interval (cf. Fig. 3), which also meets certain conditions as stated in its construction (see Section II Subsection B).

This will enable us to form a better prediction of poverty rate data than if we used Newton's interpolation polynomial. To clearly determine the superiority of a cubic spline over Newton interpolation, a comparison of both graphs will be provided in the following section.

C. Comparison of Newton's Interpolation Polynomial and a Cubic Spline

By plotting the graphs of both Newton's interpolation polynomial and a cubic spline in one figure, as shown in Fig. 6, we are able to clearly see the differences and similarities between the two sets of results. Both methods produce smooth functions passing through each data point. However, the former has greater errors than the latter through the



interval. Most notably, considerable errors occur within subintervals [2010, 2011] and [2011,2012] for the left end interval, and [2018,2019] and [2019,2020] for the other interval. Predicted data pertaining to those subintervals are certainly less accurate.

The fact that the latter has smaller errors than the former is due to the fact that the latter is constructed to be smooth while preserving the pattern of the data (see Fig. 1). Predicted data using a cubic spline can hence be more reliable than those of Newton's interpolation polynomial. Therefore, a cubic spline is superior to Newton's interpolation polynomial.

D. Newton Extrapolation and Cubic Spline Extrapolation

Newton extrapolation uses the last three years of data – namely, the data taken from 2018, 2019, and 2020 - to predict data in 2021. This is because these data bear the greatest similarity to the predicted data. With three data samples, we are able to use a quadratic polynomial as follows:

$$P_2(x) = 7.45 + (x - 2018)(-0.54 + 0.755(x - 2019)) \quad (11)$$

The plot of $P_2(x)$ is then shown in Fig. 7. Evaluating (11) at x = 2021 gives the value of 10.36 (see the arrow).

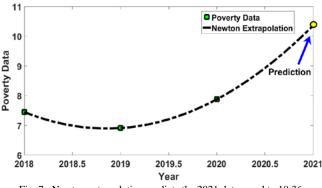


Fig. 7. Newton extrapolation predicts the 2021 data equal to 10.36.

On the other hand, cubic spline extrapolation uses the last spline function (10) to predict the 2021 data. The extrapolation is stated as follows:

(12)

$$\xi_{10}(x) = -619.2455 + 0.3101x + 0.9898(x - 2019)^{2} + -0.3299(x - 2019)^{3}$$

Fig.8 shows the graph of the extrapolation. Evaluating (12) at x = 2021 gives the value of 10.6 (see the arrow).

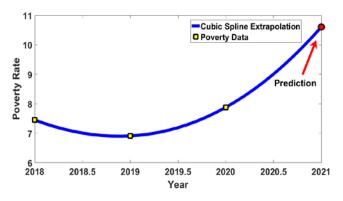


Fig. 8. Cubic spline extrapolation predicts the 2021 data equal to 10.6.

It can be seen that different scales of the vertical axis have been used in both images. Therefore, to form an appropriate comparison of both results, we plot them together in one figure (see Fig. 9).

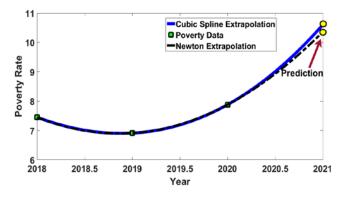


Fig. 9. The real prediction of 2021 data using cubic spline extrapolation and Newton extrapolation.

Fig. 9 shows the real prediction of the 2021 data using cubic spline extrapolation and Newton extrapolation equal to 10.6% and 10.36% respectively (see the arrow). The former, which is cubic, has a slightly higher value than the latter, which is quadratic. Furthermore, in line with the previous results (see Section III Subsection C), these results imply that the former fits the trend of the poverty data better than the latter. Therefore, cubic spline extrapolation is superior to Newton extrapolation for predicting data beyond the range of available data.

IV. CONCLUSION

In this paper, interpolation and extrapolation on poverty rate data in West Java are performed by using Newton's divided differences and a cubic spline. Our results demonstrate that a cubic spline is superior to Newton's divided differences, meaning a cubic spline can provide a better prediction of poverty rate data than Newton's divided differences both within and beyond the range of the given data.

REFERENCES

- J. A. Patel, F. B. H. Nielsen, A. A. Badiani, S. Assi, V. A. Unadkat, B. Patel, R. Ravindrane and H. Wardle, "Poverty, inequality and COVID-19: the forgotten vulnerable," *Public health* 183 (2020) 110.
- [2] A. C. Aitken, "On interpolation by iteration of proportional parts, without the use of differences," Proceedings of the Edinburgh Mathematical Society 3(1) (1932) 56-76.
- [3] J. T. Schelten and F. Hossfeld, "Application of spline functions to the correction of resolution errors in small-angle scattering," *Journal of Applied Crystallography* 4(3) (1971) 210-223.
- [4] I. A. Blatov, A. I. Zadorin, and E. V. Kitaeva, "Approximation of a Function and Its Derivatives on the Basis of Cubic Spline Interpolation in the Presence of a Boundary Layer," *Computational Mathematics and Mathematical Physics* 59 (2019) 343–354.
- [5] J. Lian, W. Yu, K. Xiao, and W. Liu, "Cubic Spline Interpolation-Based Robot Path Planning Using a Chaotic Adaptive Particle Swarm Optimization Algorithm," *Mathematical Problems in Engineering* 2020 (2020) 1-20.
- [6] J. A. Ligget and J. R. Salmon, "Cubic spline boundary elements," *International Journal for Numerical Methods in Engineering* 17(4) (1981) 543-556.
- [7] G. Behforooz, "End conditions for cubic spline interpolation derived from integration," *Applied Mathematics and Computation* 29(3) (1989) 231-244.
- [8] L. A. Dontchev, H. D. Qi, L. Qi and H. Yin, "A Newton Method for Shape-Preserving Spline Interpolation," *SIAM Journal on Optimization* 13(2) (2002) 588-602.
- [9] A. Fischer, "A Newton-type method for positive-semidefinite linear complementarity problems," *Journal of Optimization Theory and Applications* 86 (1995): 585-608.
- [10] J. Hoschek and U. Schwanecke, "Interpolation and approximation with ruled surfaces," *The mathematics of surfaces* 8 (1998) 213-231.
- [11] R. Goonatilake and V. Ruiz, "On polynomial interpolation approximations: Numerical vs. statistical techniques," *Journal of Modern Methods in Numerical Mathematics* 5(1) (2014) 17-27.
- [12] M. Gasca and T. Sauer, "Polynomial interpolation in several variables," Advances in Computational Mathematics 12 (2000) 377-410.
- [13] M. I. Syam, "Cubic spline interpolation predictors over implicitly defined curves," *Journal of Computational and Applied Mathematics*, 157(2), (2003), 283-295.
- [14] K. Atkinson and W. Han, *Elementary Numerical Analysis*, 3rd ed. New Jersey: John Wiley & Sons, Inc., 2004.
- [15] X. Qin, L. Qin, and Q. Xu, "C1 Positivity-preserving Interpolation Schemes with Local Free Parameters," IAENG International Journal of Computer Science, vol. 43, no. 2, pp219–227, 2016.
- [16] S. McKinley, & M. Levine, "Cubic spline interpolation," College of the Redwoods 45(1) (1998) 1049-1060.
- [17] Devi Munandar, Wahyu Suryaningrat, and Sri Purwani, "Polynomial Interpolation and Cubic Spline to Determine Approximate Function of COVID-19 Tweet Dataset," IAENG International Journal of Applied Mathematics, vol. 52, no. 1, pp35-43, 2022.