THE 1ST LEWIS
The 1st Lawang Sewu International Symposium
Volume 1, November 2022

E-ISSN : 2987 - 2685
page 167 - 175

Covid-19 Daily Case Modelling using Nonparametric Regression Hybrid Model

Tiani Wahyu Utami[1*], Fatkhurokhman Fauzi[1], Noor Wahida Binti Md. Junos[2]
[1]Faculty of Science and Mathematics, Universitas Muhammadiyah Semarang, Indonesia
[2]Faculty of Science and Mathematics, Univesiti Pendidikan Sultan Idris, Malaysia
[*]Corresponding author : tianiutami@unimus.ac.id

ABSTRACT

Nonparametric Regression Model is used to analyze data patterns and predict the future based on past data. The problem of Nonparametric Regression research so far has only used one approach, causing the estimation results to be biased, even though each sub-pattern of data has its own suitability depending on the approach method used. In this study, a hybrid nonparametric regression modeling model based on two approaches will be developed. The approach used in this research is Local Polynomial and Fourier Series. Coronaviruses are part of the family of viruses caused by the corona virus, otherwise known as COVID-19, which is a new type that was discovered in 2019. The spread of the corona virus in Semarang City continues to increase dramatically. Therefore, the purpose of this study is to estimate the Local Polynomial- Fourier Series Hybrid model which is then applied to the Covid-19 case in Semarang City. The following is the research plan or research stages for the Primary Research Scheme, namely estimating the Hybrid model with two Local Polynomials and Fourier Series approaches, the estimation method used is Weighted Least Square (WLS). In the Local Polynomial- Fourier Series hybrid model there is a smoothing parameter selection, the method used is the Generalized Cross Validation (GCV) method. Furthermore, the program that has been created is applied to Covid-19 data in Central Java so as to produce Covid-19 case modeling based on Hybrid Nonparametric Regression using two approaches, namely Local Polynomials and Fourier Series. The processed results using nonparametric regression hybrid model of local polynomial - fourier series approach applied to daily case data of Novel Coronavirus in Semarang City showed a coefficient of determination of 0.9968 (99.68%) and an MSE of 313.153.

Keywords: Covid-19, Model Hybrid, Local Polynomial, Nonparametric Regression

1. INTRODUCTION

Regression analysis was conducted to determine the relationship and influence of the predictor variables on the response variables by estimating the regression curve function. We can model with a flexible statistical approach without any assumptions like the nonparametric regression model. Nonparametric Regression Model is used to analyze data patterns and predict the future based on past data. Some nonparametric regression approaches that are often used include Truncated Spline, Fourier Series, Kernel Local Polynomials, Wavelets and so on [1]. Nonparametric regression modeling develops every period. The problem of Nonparametric Regression research so far has only used one approach, causing the estimation results to be biased, even though each sub-pattern of data has its own suitability depending on the approach method used. Therefore, the Hybrid

E-ISSN : 2987 - 2685
page 167 - 175

method emerged as a development of Nonparametric Regression. Hybrid model is a combined model between approach methods, with the hope of increasing accuracy in modeling analysis. The nonparametric regression approach in this study was carried out using two approaches, namely Local Polynomials-Fourier Series.

At the beginning of 2020 the world was shocked by the new Corona virus outbreak (Covid-19) which infected almost all countries in the world. The World Health Organization has declared a global emergency due to the virus since January 2020 [2]. At the beginning of Covid-19 in Indonesia the government has taken various policy initiatives against the virus starting with health social economy and other sectors. Coronavirus disease (COVID 19) is an infectious disease caused by the recently identified coronavirus 2 (SARS-CoV-2) and severe acute respiratory syndrome or known as coronavirus. COVID 19 has become an epidemic that is now happening in many countries around the world and Indonesia is one of them. Central Java is one of the regions in Indonesia where many cases of Covid-19 have been detected [3]. Semarang is the city with the highest number of COVID-19 cases in Central Java. Data on COVID cases in Semarang City does not form a specific distribution pattern. The choice of Semarang City was based on the fact that Semarang City has been included in one of the Covid-19 Red Zones since it was established at the end of February 2020. In addition, Semarang City is also a link between West Java and East Java via the North route. The Coastline (Pantura) which is traversed by many vehicles and is one of the main routes for domestic and international flights in Central Java Province.

1.1. The Fourier Series

Fourier series are flexible trigonometric polynomial functions. This is because the Fourier series is the curve representing the cosine of the sine function [4]. The Fourier Series function is as follows:

$$f(x) = \frac{1}{2}a_0 + \gamma x + \sum_{k=1}^{K} a_k \cos(\frac{2\pi kx}{2L})$$

The smoothness level of the Fourier series estimator is determined by the selection of the smoothing parameter K. The smaller the smoothing parameter K, the smoother the estimation and the larger the smoothing parameter K, the less smooth the estimation of f [5]. Therefore, it is necessary to choose the optimal K.

1.2. Local Polynomial Nonparametric Regression

Nonparametric regression is a method used to estimate the pattern of the relationship between the response variable and the predictor variable, where the shape of the regression curve is unknown [6]. Given data $(x_i, y_i), i = 1,2,..n$ where *n* is the number of subjects. The x_i is the predictor variable observed from the i^{th} subject. The relationship between these variables is stated in the nonparametric regression model as follows:

$$y_i = ç(x_i) + e_i; i = 1,2,..n \quad (1)$$

The function $\eta(x_i)$ is a function that has no known form called a regression function [7]. Where $e_i \square N(0, \sigma^2)$ is the measurement error. It is known that x is the predictor

E-ISSN : 2987 - 2685
page 167 - 175

variable so that the function η is estimated using the Local Polynomial Kernel approach. With the Taylor series, $\eta(x_i)$ in equation (1) can be approximated by a polynomial of degree p as follows:

$$\eta(x_i) \approx \eta(x) + (x_i - x)\eta^{(1)}(x) + \cdots + (x_i - x)^p \eta^{(p)}(x)/p!$$

$$x_i \in [x - h, x + h]$$

Suppose $\beta_r(x) = \eta^{(r)}(x)/r!; r = 0,1,2,\ldots,p$ then can be written as:$\eta(x_i) \approx \beta_0(x) + (x_i - x)\beta_1(x) + \cdots + (x_i - x)^p \beta_p(x)$

1.3. Corona Virus Disease 19 (Covid-19)

Corona Virus Disease 19 (Covid-19) was first known at the end of 2019 at Wuhan city of China. This virus attacks the respiratory system with symptoms such as pneumonia. This virus is a relatively new virus, so it doesn't have one antidote and this virus has spread all over the world until it gets out of control. It has been recorded that more than 200 countries have reported cases of COVID-19, including Indonesia [3]. In the current conditions the corona virus has been established by [8] as a pandemic. The spread of Covid-19 has entered Indonesia since March 2, 2020.

Indonesia a developing country and the fourth most populous country in the world is at high risk and is expected to face the threat of COVID-19 harder and longer than other countries. [9]. The Central Java Provincial Government (Pemprov) records the number of the latest corona virus cases on Tuesday, July 28, 2020 at 12.00 WIB. Through its official website corona.jatengprov.go.id, the Central Java Provincial Government provides information on corona virus case data. The site recorded 8,795 people infected with the corona virus, and 4,902 people had been declared cured. In the Central Java region there were also 750 patients who died from this virus. So that as many as 3,143 people are still being treated at the hospital designated to treat corona virus patients. There are also 546 people in the monitoring site (ODP) who are currently in the monitoring process in Central Java.

2. METHOD

The main problem to be solved in this research is the development and application of the Hybrid Nonparametric Regression model on daily Covid-19 case data in Semarang City and the development of an opensource program using R Software. This study uses two approach methods, namely Local Polynomials-Fourier Series. The following covers several stages of research.

Nonparametric Regression Hybrid Modeling using Local Polynomials-Fourier Series on Covid-19 data:

a. Estimation of Local Polynomial Nonparametric Regression Model using the WLS method;
b. Create a program to determine the optimum bandwidth with the GCV method;
c. Use optimum bandwidth to determine Kernel regression parameter estimation program;
d. Calculate the residual value of the Local Polynomial regression model;

E-ISSN : 2987 - 2685
page 167 - 175

e. Estimated Fourier Series with the response variable is the residual value of the Local Polynomial regression model;
f. Create a program to determine the optimum K in the Fourier Series model;
g. Use the optimum K to determine the parameter estimates for the Fourier Series regression model;
h. Calculate the MSE and R-square values of the Hybrid model. Selection of the best model based on the smallest MSE value and the largest R-square.

3. RESULTS AND DISCUSSION

3.1. Modeling of Covid-19 Cases in Semarang City with a Hybrid Regression Approach to Local Polynomials-Fourier Series

Coronavirus disease 2019 (COVID 19) is an infectious disease caused by acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or a newly discovered coronavirus. Covid 19 is now spreading like an epidemic in many countries around the world including Indonesia. Central Java Indonesia is the region with the highest number of registered cases of coronavirus infection. The city of Semarang has the highest number of his Covid-19 cases in Central Java for the first time. From April 9 to August 7, 2020, 528 Covid-19 cases occurred in Semarang City.

The data of positive covid patients in Semarang does not show a specific distribution pattern. We can build models without any assumptions with flexible statistical procedures especially non-parametric regression procedures. A non-parametric programming approach using local polynomials was used in this study. In Local Polynomial Regression modeling, the determination of the order of the polynomial and the optimal bandwidth uses the GCV method. In the first modeling, the data used is data on the number of positive patients for the Coronavirus or COVID 19 in Semarang City, Indonesia using the Kernel Local Polynomial approach. The city of Semarang was chosen as the research location. The reason for choosing Semarang City is because Semarang City is one of the Covid-19 Red Zones since it was established at the end of February 2020. In addition, Semarang City is also a link between West Java and East Java via the northern route. The Coastal Line (Pantura) which is widely used by vehicles and is one of the main routes for domestic and international flights in Central Java Province.

The data used for the implementation of the local polynomial kernel-fourier series nonparametric hybrid regression model is the daily number of corona virus cases. Before getting a nonparametric hybrid regression modeling using the local polynomial kernel approach - Fourier series is to create a scatterplot of Covid case data in Semarang City. The following is a scatterplot of daily cases of the Corona virus in Semarang City, Central Java with observations from April 9, 2020 to August 7, 2020:

E-ISSN : 2987 - 2685
page 167 - 175

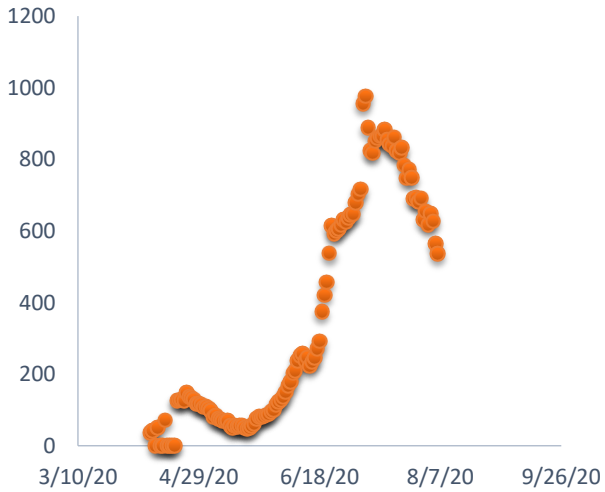

Figure 1. *Scatterplot graph of daily cases of Corona virus in Semarang City, Central Java*

Based on the scatterplot graph of the Novel Coronavirus Daily Cases, it can be said that with the number of cases of COVID in Semarang city Figure 1 shows that it does not form a specific pattern so that the modeling is approached by nonparametric regression with a local polynomial kernel-fourier series approach. The data was obtained from the Semarang City Corona Alert 2020 with daily observations from 9 April 2020 to 7 August 2020. Semarang City had the highest number of COVID cases in early July 2020. The response variable for this study is the number of Covid cases. While the predictor variable in this study is the time of observation. Data on COVID cases in Semarang City does not form a specific distribution pattern.

3.2. Estimating Nonparametric Regression Hybrid Models of Fourier Series Based on Residual Values of Kernel Local Polynomial Models

Determining the model with the Fourier series approach based on the residual value of the local polynomial model by finding the optimal K value by running the optimal K determination program that was previously made on the daily Covid-19 data in Semarang City. The value of K is a positive integer. Determination of the optimal K value in this study using the GCV method. The results obtained from K were tested on the residual data of the local polynomial model as follows.

Based on Table 1, the GCV value at K = 121 is the most optimal K because the GCV value is the smallest among other GCV values. If the value of K = 121 is used, the number of parameters that must be estimated is 123 parameters. Furthermore, the determination of the best model can be seen from the R^2 and MSE values for each K value. The following are the R^2 and MSE values for K=85 to K= 121:

E-ISSN : 2987 - 2685
page 167 - 175

Table 1*. GCV Data value for each K is optimal*

Nilai K	GCV	Nilai K	GCV	Nilai K	GCV
81	$1{,}061\times 10^{+03}$	112	$4{,}693\times 10^{+01}$	117	9,3933
86	$5{,}438\times 10^{+02}$	113	$3{,}959\times 10^{+01}$	118	$2{,}735\times 10^{-01}$
89	$4{,}531\times 10^{+02}$	114	$2{,}248\times 10^{+01}$	119	$4{,}539\times 10^{-15}$
90	$4{,}373\times 10^{+02}$	115	$1{,}6\times 10^{+01}$	120	1×10^{-22}
95	$3{,}319\times 10^{+02}$	116	$1{,}573\times 10^{+01}$	121	$9{,}096\times 10^{-24}$

Table 2. *R^2 and MSE Values Every K is Optimal*

Nilai K	R^2	MSE
85	0,904 (90,4%)	281,1479
90	0,9154(91,54%)	247,3649
105	0,9764 (97,64 %)	69,1448
109	0,9831 (98,31 %)	49,3396
115	0,9950 (99,5 %)	14,7108
118	0,9999(99,99%)	0,2645

From the Table 2, it can be seen that the value of K = 115 has produced a fairly high R2, while for the value of K = 90 it has produced an R^2 of 91.54%. If the selected K value is K = 85, then the estimated parameters that must be searched are 88 parameters, this can be seen in Equation 5.2 by looking at the number of estimated parameters. The criteria for selecting the model are a model with a large R^2, a small MSE value and a parsimony (simple) model, so that the selected model is a model that has an optimal K value of 85.

After the estimation of the local kernel polynomial model (t_i) and the estimation results of the model using the Fourier series approach f(e) are obtained, the two models are combined so that they become a Nonparametric Regression model for Kernel-Fourier Local Kernel-Fourier Series Hybrid Nonparametric Polynomials. So that the Kernel-Fourier Series Nonparametric Hybrid Nonparametric Regression model becomes as follows:

$$\hat{y}_i = \eta(x_i) + \frac{1}{2}a_0 + \gamma x + \sum_{k=1}^{K} a_k \cos\left(\frac{2\pi k \varepsilon_i}{2L}\right) + e_i$$

The following are the estimation results of the Kernel-Fourier Series Local Polynomial Hybrid Nonparametric Regression model which is applied to the daily Covid-19 cases in the City of Semarang which is presented together with the actual data on the daily Covid cases in the City of Semarang :

E-ISSN : 2987 - 2685
page 167 - 175

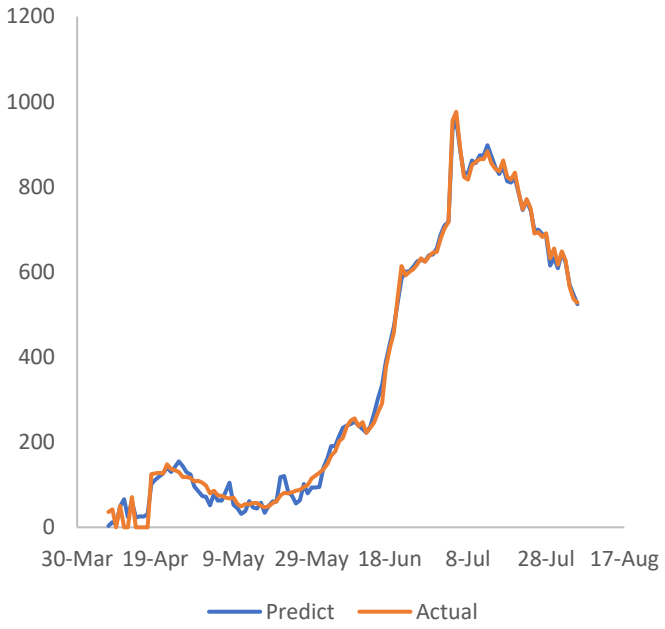

Figure 2. *Graph of Local Polynomial-Fourier Series Nonparametric Regression Hybrid Model*

Based on Figure 2, it can be concluded that the dynamics of changing patterns between the actual data and the same data. It can be said that the modeling obtained is suitable for predicting cases of the Corona virus in the city of Semarang. The calculation results show the total number of corona cases in Semarang city at the beginning of July 2020. After the corona cases increased in July while the corona cases decreased in August.

The ability of the model to make accurate predictions can be demonstrated by the coefficient of determination (R^2) values and the mean square error (MSE) values. The model is said to be better if the determination coefficient is close to 100% and MSE value is getting smaller. The processed results using a nonparametric hybrid polynomial local kernel-fourier series regression approach applied to daily case data of Novel Coronavirus in Semarang City showed a coefficient of determination of 0.9968 (99.68%) and an MSE of 313.153. The coefficient of determination of 99.68% means that the predictor variable, namely the time of observation, can explain the number of daily cases of COVID patients in Semarang City by 99.68%, while 0.32% is influenced by other variables not used in this study. The following table compares the Local Polynomial model with the Local Polynomial-Fourier Series hybrid model:

Based on Table 3, it can be compared that the Local Polynomial-Fourier Series hybrid model is better than the Local Polynomial model, it can be seen from the largest R-square, which is 99.68% and the smallest MSE. It can be concluded that the existence of a hybrid model can further minimize the error value so that the resulting R-square is close to 100%.

E-ISSN : 2987 - 2685
page 167 - 175

Table 3. *Comparison of Local Polynomial Model with Local Polynomial-Fourier Series Hybrid Model*

Nonparametric Regression	R-Square (%)	MSE
Local Polynomial	97	2926,669
Local Polynomial-Fourier Series Hybrid Model	**99,68**	**313,153**

The dynamics of changing patterns between the actual data and the predicted data are the same. It can be said that the modeling obtained is suitable for predicting cases of the Corona virus in the city of Semarang. The Estimated results show that Semarang City recorded the highest number of corona virus cases in early July 2020. Corona virus cases decreased in August after increasing in July. The predictive ability of the model can be seen in the correlation coefficient (R^2) and the Mean Square Error (MSE). A model is said to be better if the coefficient of determination is close to 100% and the MSE value is getting smaller.

4. CONCLUSION

The Kernel-Fourier Series Nonparametric Hybrid Nonparametric Regression model becomes as follows:

$$\hat{y}_i = \eta(x_i) + \frac{1}{2}a_0 + \gamma x + \sum_{k=1}^{K} a_k \cos(\frac{2\pi k \varepsilon_i}{2L}) + e_i$$

The processed results using nonparametric regression hybrid model of local polynomial - fourier series approach applied to daily case data of Novel Coronavirus in Semarang City showed a coefficient of determination of 0.9968 (99.68%) and an MSE of 313.153.

Based on the results obtained, it can be compared that the hybrid model of the Local Polynomial-Fourier Series is better than the Local Polynomial model, this can be seen from the largest R-square, which is 99.68% and the smallest MSE. Then from the results of the study it can be concluded that the existence of a hybrid model can further minimize the error value so that the resulting R-square is close to 100%.

5. ACKNOWLEDGMENTS

The author would like to thanks Universitas Muhammadiyah Semarang's Institute for Research and Community Services for funding this researchment, as well as the Mathematics and Sciences' Laboratory which has facilitated the data processing.

REFERENCES

[1] D. Kong, H. D. Bondell, and Y. Wu, "Domain selection for the varying coefficient model via local polynomial regression," *Comput Stat Data Anal*, vol. 83, pp. 236–250, 2015, doi: https://doi.org/10.1016/j.csda.2014.10.004.

E-ISSN : 2987 - 2685
page 167 - 175

[2] H. Nugroho, "Indonesia Development Update A Year of Covid-19: A Long Road to Recovery and Acceleration of Indonesia's Development," *Jurnal Perencanaan Pembangunan The Indonesian Journal of Development Planning*, vol. V, no. 1, doi: 10.36574/jpp.v5i1.

[3] D. Handayani, D. H. Hadi, F. Sbaniah, E. Burhan, and H. Agustin, "Penyakit Virus Corona 2019," *Jurnal Respirologi Indonesia*, vol. 4, no. 2, pp. 119–129, 2020.

[4] R. Nur Wisisono, A. I. Nurwahidah, and Y. Andriyana, "Regresi Nonparametrik dengan Pendekatan Deret Fourier pada Data Debit Air Sungai Citarum," *Jurnal Matematika "MANTIK,"* vol. 4, no. 2, pp. 75–82, Oct. 2018, doi: 10.15642/mantik.2018.4.2.75-82.

[5] T. W. Utami, M. A. Haris, A. Prahutama, and E. A. Purnomo, "Optimal knot selection in spline regression using unbiased risk and generalized cross validation methods," in *Journal of Physics: Conference Series*, Jan. 2020, vol. 1446, no. 1. doi: 10.1088/1742-6596/1446/1/012049.

[6] T. W. Utami, A. Prahutama, A. Karim, and A. R. F. Achmad, "Modelling rice production in Central Java using semiparametric regression of local polynomial kernel approach," in *Journal of Physics: Conference Series*, Jun. 2019, vol. 1217, no. 1. doi: 10.1088/1742-6596/1217/1/012108.

[7] T. W. Utami, "ESTIMASI KURVA REGRESI SEMIPARAMETRIK PADA DATA LONGITUDINAL BERDASARKAN ESTIMATOR POLINOMIAL LOKAL," *Jurnal Statistika Universitas Muhammadiyah Semarang*, vol. 1, no. 1, pp. 30–36, 2013, [Online]. Available: http://jurnal.unimus.ac.id.

[8] World Health Organization, "Situation Report-51 SITUATION IN NUMBERS total and new cases in last 24 hours," 2020. Accessed: Dec. 19, 2022. [Online]. Available: https://apps.who.int/iris/handle/10665/331475.

[9] R. Djalante *et al.*, "Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020," *Progress in Disaster Science*, vol. 6, p. 100091, 2020, doi: https://doi.org/10.1016/j.pdisas.2020.100091.