KERNEL NONPARAMETRIC REGRESSION FOR THE MODELIZING OF THE PRODUCTIVITY WETLAND PADDY

Tiani Wahyu Utami[1], Martyana Prihaswati [2], Vega Zayu Varima[3]

[1,2,3]Universitas Muhammadiyah Semarang
[1]tianiutami@unimus.ac.id

Abstract

Nonparametric regression can be used when the relationship between the response variable and the predictor variables have an unknown pattern form the regression curve. One of the method that can be used to predict productivity of the wetland paddy is a nonparametric regression kernel. In kernel regression, there are several types of estimator that can be used to modelling productivity of wetland paddy in Central Java, one of which is Nadaraya-Watson estimator. Variables used in the study of the productivity of rice as the response variable, while the predictor variables that harvested area, production and rainfall. Based on estimates indicate that the kernel nonparametric regression optimum bandwidth value 1.2 and GCV = 1.7577. The coefficient of determination (R^2) of 94.23% and MSE of 0.8560.

Keywords: Kernel Nonparametric Regression, Productivity, Wetland Paddy

1. Introduction

In regression analysis, there are two approaches, parametric approach and nonparametric approach. A parametric regression model requires an assumption that the form of the underlying regression function is known except for the value of a finite number of parameters. The selection of parametric model depends very much on the problem at hand. A serious drawback of parametric modeling is that a parametric model may be too restrictive in some applications. If an in appropriate parametric model is used, it is possible to produce misleading conclusions from the regression analysis. In other situations, a parametric model may not be available to use. To overcome the difficulty caused by restrictive assumption of a parametric from of the regression function, one may remove the restriction that the regression function belongs to a parametric family. This approach leads to so-called nonparametric regression [6].

There exist many nonparametric regression and smoothing method. One of methods used in this writing is by regression nonparametic approaching, because the approach does not depend on the assumption of a certain shape of the curve, thus providing greater flexibility [5]. The most popular methods include kernel, spline, local polynomial kernel, and deret fourier. One of regression nonparametric approaching used in this writing is Kernel. The strength of Kernel of regression nonparametric approaching is that it enables to solve easy and simple. In regression nonparametric kernel, there are several types of estimator that can be used to modelling, one of which is Nadaraya-Watson estimator.

Researchs about regression nonparametic approaching of Kernel were done previously by [4] about Nadaraya-Watson estimator is method better of estimating model based on approaching which is not tied to the assumption in the form of certain regression curva, and research about bandwidth optimum with kernel regression semiparametrik approaching by [3].

Indonesia Ministry of Agriculture data indicates that from 19981, to 1985 and from 1998 to 1999, the conversion of paddy State of Indonesia is an agricultural country with one of the staple food in Indonesia is rice [1]. The majority of rice field conversion took place in Java, which had accounted for 60% of national rice production. Production of rice in the province of Central Java is the highest in Java. In Central Java according to the Central Bureau of Statistics paddy crop production is fluctuating [2].

This results in the need for modeling to predict and know how change productivity wetland paddy in Central Java. The result of the modeling is expected to help the concerning parties the strategical steps is needed to be done so that not suffering significant losses. Therefore, the appropriate statistical method for modeling productivity wetland paddy in Central Java is using the nonparametric regression approach of Kernel.

The method of kernel nonparametric regression is the regression method used when the curva is between dependent and independent variable, and Independent variable is not known for

the form and pattern. The common nonparametric regression model is as follows:

$$y_i = f(x_i) + \varepsilon_i \quad (1)$$

y_i = dependent variable
x_i = independent variable
$f(x_i)$ = regression function

2. Method

2.1 Data Resources

The main data resourses used in this research is the secondary data served by the the Central Bureau of Statistics (BPS). The data in this study is a secondary data, in the District / City in Central Java at 2015 were obtained from the Central Bureau of Statistics (BPS) with the response variable is productivity wetland paddy.

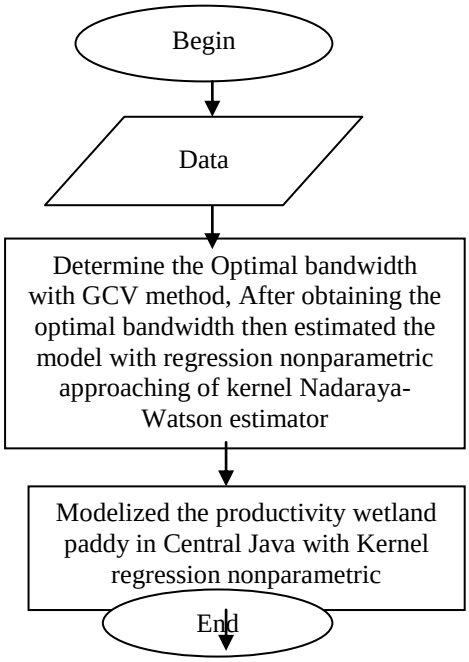
2.2 Research Variable

Table 1. Research Variable

Data	Variabel	Variable Information
Produktivity	Respon	Respon
Harvested area	Predictor	Nonparametrik
Production	Prediktor	Nonparametrik
Rainfall	Prediktor	Nonparametrik

2.3 Procedures (or research design)

Analysis steps in this research can be described in diagram as described in Flowchart **2.1** below:



Flowchart 1. Analysis steps in this research

3.Results

3.1 Determine Optimum Bandwidth

The first step before estimating the regression model is to determine the optimum bandwidth, namely bandwidth which has a minimum value of GCV. The optimum bandwidth *h* value is a positive integer. The determination of the optimum bandwidth *h* value is using GCV method then running the program of the determination of optimal bandwidth (*h)* value on productivity wetland paddy in Central Java based on GCV method. The obtained result from the tested bandwidth (*h)* is as follows:

Table 2. The Value Using GCV Method to Every bandwidth (*h)*

Bandwidth (*h)*	GCV	MSE
1,2	**1,7577**	**0,8560**
1,3	1,9525	0,9583
1,4	2,1860	1,0813
1,5	2,4583	1,2254
1,6	2,7651	1,3889
1,7	3,0992	1,5687
1,8	3,4530	1,7612
1,9	3,8191	1,9626
2	4,5630	2,3797

Table 2 shows that the optimum bandwidth (*h)* on the average data of productivity wetland paddy in Central Java is on h=1.2 because of the lowest GCV value. Then the optimum bandwidth is used to estimate the smoothing function. MSE obtained amounted to 0.8560 viewed from the optimal bandwidth minimum of 1.2 GCV at 1.7577. Based on **Table 2,** it shows that for the optimum bandwidth = 1.2 has resulted R^2 =94.23 % which is enough high.

3.2 The Modelizing of Productivity Wetland Paddy with Regression Nonparametric Kernel

After knowing that the optimum bandwidth is 1.2, the next step is to determine the estimation model of productivity wetland paddy with regression nonparametric approaching of Kernel. The result of estimated model can be seen on equation 2. Equation 2 shows that the obtained model for productivity wetland paddy data in Cental Java as follows :

$$\hat{y} = \frac{\sum_{i=1}^{n} K_{1.2}\left(\frac{x - x_{1i}}{1.2}\right) y_i}{\sum_{i=1}^{n} K_{1.2}\left(\frac{x - x_{1i}}{1.2}\right)} + \frac{\sum_{i=1}^{n} K_{1.2}\left(\frac{x - x_{2i}}{1.2}\right) y_i}{\sum_{i=1}^{n} K_{1.2}\left(\frac{x - x_{2i}}{1.2}\right)} + \frac{\sum_{i=1}^{n} K_{1.2}\left(\frac{x - x_{3i}}{1.2}\right) y_i}{\sum_{i=1}^{n} K_{1.2}\left(\frac{x - x_{3i}}{1.2}\right)} \quad (2)$$

The plot of estimation results of productivity wetland paddy use an optimal bandwidth can be seen in Figure 1 as follows :

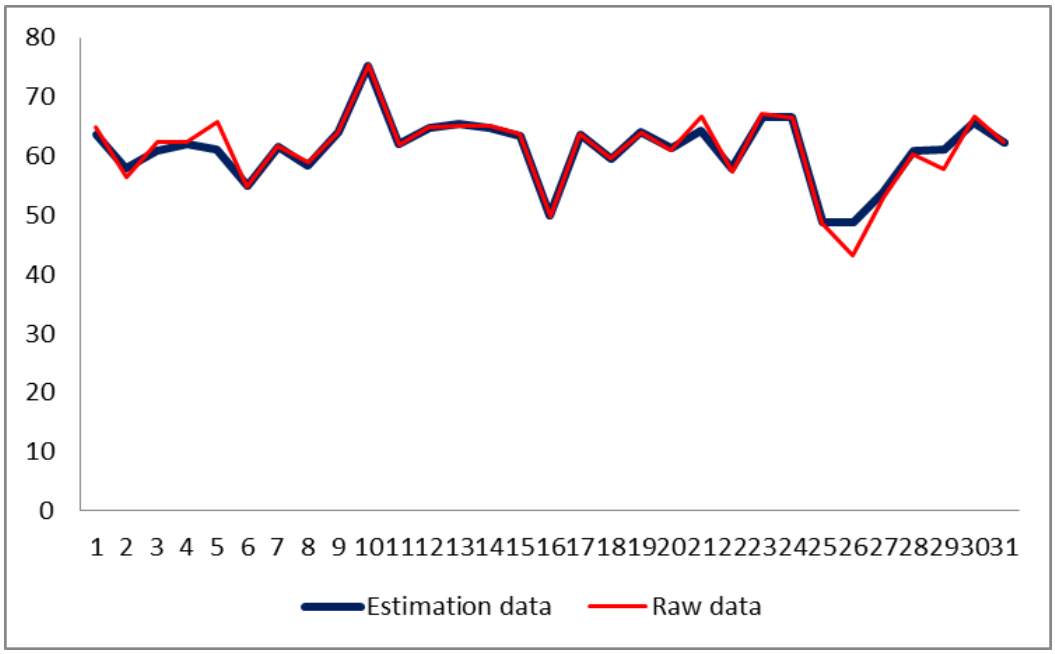


Figure 1. Plot Estimation Productivity Wetland Paddy Using Nonparametric Regression Kernel

4. Discussion

Based on Figure 2, it can be seen that raw data with estimation data have pattern is same. The estimation result of the lowest productivity wetland paddy in Central Java in Pekalongan. The estimation result of highest productivity wetland paddy in Central Java in Klaten. The result of the model can be used to forecast the average productivity wetland paddy that will be going to happen in the future by entering value variable that can be predicted in the equation 2.

5. Acknowledgement

Thanks to Direktorat Riset dan Pengabdian Masyarakat Direktorat Jenderal Penguatan Riset dan Pengembangan Kementrian Riset, Teknologi, dan Pendidikan Tinggi Penelitian Dosen Pemula.

REFERENCES

1. Balai Besar Penelitian Tanaman Padi. (2015). BPS: Produksi Padi Tertinggi Dalam 10 Tahun Terakhir. "http://bbpadi.litbang.pertanian.go.id/index.php/berita/beritautama/content/106-bps-produksi-padi-tertinggi-dalam-10-tahun-terakhir"
2. BPS (2015). "Data BPS Provinsi Jawa Timu". BPS Jawa Timur.
3. Budiantara, I.N. dan Mulianah (2007). "Pemilihan Banwidth Optimal Dalam Regresi Semiparametrik Kernel dan Aplikasinya", Journal Sains dan Teknologi SIGMA, 10 : 159-166.
4. Esty (2014). "Regresi Kernel Dengan Metode Nadaraya Watson". Skripsi. Universitas Negeri Yogyakarta. Yogyakarta

5. Utami, T. W and Nur, I. M (2014). "Longitudinal Data Modeling Based On Platelet Level in DHF (Dengue Hemorrhagic Fever) Using Nonparametric Regression of Local Polynomial Kernel GEE Approach", International Conference on Biomedical Engineering Technology and Application. UGM, Yogyakarta.
6. Wu, H. and Zhang (2006)."Nonparametric Regression Methods for Longitudinal Data Analysis", John Wiley and Sons, Inc., New York.