

ABSTRAK

Regresi *spline* merupakan salah satu model regresi non parametrik yang digunakan untuk mengestimasi pola hubungan pasangan data variabel prediktor dengan variabel respon dengan tetap memperhatikan kemulusan kurva. Pada regresi *spline*, penentuan banyaknya titik knot dan lokasi titik knot dapat mempengaruhi nilai error dari model. Tujuan dari penelitian ini adalah untuk mendapatkan model regresi *spline* linier multivariabel terbaik menggunakan kriteria *generalized cross validation*. Data yang digunakan dalam penelitian ini adalah data tingkat kemiskinan di Kabupaten Banyumas tahun 2015. Hasil penelitian menunjukkan bahwa PDRB perkapita di Kabupaten Banyumas (X_1) dan presentase pendidikan rendah di Kabupaten Banyumas (X_5) merupakan faktor yang berpengaruh terhadap presentase tingkat kemiskinan di Kabupaten Banyumas (Y). Estimasi model regresi *spline* linier multivariabel terbaik yang diperoleh yaitu:

$$\begin{aligned}\hat{Y}(x) = & -38,9486631 - 6,1953640x_1 + 22,0448736(x_1 - 7,223)_+ \\ & - 37,9794971(x_1 - 8,274)_+ + 28,0903573(x_1 - 9,017)_+ \\ & - 21,0921897(x_1 - 11,093)_+ + 15,8888925(x_1 - 11,766)_+ \\ & + 2,0178601x_5 - 7,0229731(x_5 - 52,07)_+ \\ & + 14,6925871(x_5 - 55,86)_+ - 28,0945231(x_5 - 58,01)_+ \\ & + 18,8118903(x_5 - 58,43)_+ - 0,1025776(x_5 - 70,00)_+\end{aligned}$$

Dengan knot $K_1 = 7,223$; $K_2 = 8,274$; $K_3 = 9,017$; $K_4 = 11,093$; $K_5 = 11,766$ berada pada variabel X_1 , sedangkan $K_6 = 52,07$; $K_7 = 55,86$; $K_8 = 58,01$; $K_9 = 58,43$ dan $K_{10} = 70,00$ berada pada variabel X_5 . Nilai GCV minimum 28,95036 serta nilai koefisien determinasi sebesar 89,84299%

Kata kunci: *generalized cross validation*, tingkat kemiskinan, regresi *spline* linier multivariabel, titik knot.

ABSTRACT

Spline regression is one of the non-parametric regression models that is used to estimate the correlation pattern between predictor variable with response variable while considering the smoothness of the curve. In *spline* regression, the selection of the number of knots and its location can affects the error value of the model. The aim of this research is to select the best multivariable linear *spline* regression model using *generalized cross validation* criteria. The data was about poverty levels in Banyumas 2015. The results showed that the PDRB perkapita in Banyumas (X_1) and percentage of low education in Banyumas (X_5) is the most influential factors on the percentage of poverty levels in Banyumas (Y). the Estimation model of the best multivariable linear *spline* regression was:

$$\begin{aligned}\hat{Y}(x) = & -38,9486631 - 6,1953640x_1 + 22,0448736(x_1 - 7,223)_+ \\ & -37,9794971(x_1 - 8,274)_+ + 28,0903573(x_1 - 9,017)_+ \\ & -21,0921897(x_1 - 11,093)_+ + 15,8888925(x_1 - 11,766)_+ \\ & + 2,0178601x_5 - 7,0229731(x_5 - 52,07)_+ \\ & + 14,6925871(x_5 - 55,86)_+ - 28,0945231(x_5 - 58,01)_+ \\ & + 18,8118903(x_5 - 58,43)_+ - 0,1025776(x_5 - 70,00)_+\end{aligned}$$

with knot $K_1 = 7,223$; $K_2 = 8,274$; $K_3 = 9,017$; $K_4 = 11,093$; $K_5 = 11,766$ are at variable X_1 , while knots $K_6 = 52,07$; $K_7 = 55,86$; $K_8 = 58,01$; $K_9 = 58,43$ dan $K_{10} = 70,00$ are at variable X_5 . The value of minimum GCV is 28,95036 and the coefficient of determination is 89,84299%.

Keywords: *generalized cross validation*, poverty level, multivariable linear *spline* regression, knots.