# REGRESI CAMPURAN NONPARAMETRIK SPLINE LINIER TRUNCATED DAN FUNGSI KERNEL UNTUK PEMODELAN DATA KEMISKINAN DI PROVINSI PAPUA


Nama Mahasiswa : Rory
NRP : 1314 201 711
Pembimbing : Prof. Dr. Drs. I Nyoman Budiantara, M.Si.
Co-Pembimbing : Dr. Wahyu Wibowo, S.Si., M.Si.


## ABSTRAK


Model regresi campuran nonparametrik $y_i = f(u_i, \tilde{v}_i) + \varepsilon_i, i = 1,2, \dots, n$, $\tilde{v}_i = (v_{1i}, v_{2i}, \dots, v_{mi})^T$ memiliki kurva regresi bersifat aditif $f(u_i, \tilde{v}_i) = g(u_i) + \sum_{j=1}^{m} h_j(v_{ji})$. Komponen $g(u_i)$ dihampiri dengan spline linier truncated, sedangkan komponen $h_j(v_{ji})$ dihampiri dengan kernel Nadaraya-Watson. Error random $\varepsilon_i$ mengikuti distribusi normal $N(0, \sigma^2)$. Tujuan dari penelitian ini adalah melakukan kajian mengenai estimator kurva regresi campuran nonparametrik spline dan kernel $f(u_i, \tilde{v}_i)$ dan mengaplikasikannya pada data kemiskinan di Provinsi Papua. Hasil kajian menunjukkan bahwa estimator kurva regresi spline $\tilde{g}(u)$ adalah $\hat{\tilde{g}}_{\tilde{\phi},\tilde{\xi}}(u, \tilde{v}) = \mathbf{S}(\tilde{\xi}, \tilde{\phi})\tilde{y}$ dan estimator kurva regresi kernel $\sum_{j=1}^{m} \tilde{h}_j(v_j)$ adalah $\sum_{j=1}^{m} \hat{\tilde{h}}_j(v_j) = \mathbf{V}(\tilde{\phi})\tilde{y}$. Selanjutnya, estimator kurva regresi campuran nonparametrik spline dan kernel $\tilde{f}(u, \tilde{v})$ adalah $\hat{\tilde{f}}_{\tilde{\phi},\tilde{\xi}}(u, \tilde{v}) = \mathbf{Z}(\tilde{\xi}, \tilde{\phi})\tilde{y}$, dimana $\mathbf{Z}(\tilde{\xi}, \tilde{\phi}) = \mathbf{S}(\tilde{\xi}, \tilde{\phi}) + \mathbf{V}(\tilde{\phi})$. Matriks $\mathbf{S}(\tilde{\xi}, \tilde{\phi})$, $\mathbf{V}(\tilde{\phi})$ dan $\mathbf{Z}(\tilde{\xi}, \tilde{\phi})$ tergantung pada lokasi titik-titik knot $\tilde{\xi}$ dan bandwidth $\tilde{\phi}$. Estimator-estimator tersebut adalah estimator bias, namun masih kelas estimator linier. Model regresi campuran nonparametrik terbaik adalah model yang menggunakan banyaknya titik knot, lokasi titik-titik knot dan bandwidth optimum yang diperoleh dengan meminimumkan fungsi *Generalized Cross Validation* (GCV). Pemilihan lokasi titik-titik knot dan bandwidth dilakukan secara simultan. Model regresi campuran nonparametrik spline dan kernel diterapkan pada data kemiskinan di Provinsi Papua, dimana sebagai variabel responnya adalah persentase penduduk miskin ($y$), variabel prediktor yang mengikuti kurva regresi spline adalah PDRB perkapita ($u$), dan variabel-variabel prediktor yang mengikuti kurva regresi kernel adalah gini ratio ($v_1$), rata-rata lama sekolah ($v_2$), tingkat pengangguran terbuka ($v_3$) dan laju pertumbuhan ekonomi ($v_4$). Model terbaik diperoleh ketika model menggunakan 3 titik knot. Estimasi model memberikan $R^2$ sebesar 92,02%. Model dapat digunakan untuk skenario kebijakan.



Kata kunci: kernel nadaraya-watson, regresi campuran nonparametrik, regresi nonparametrik aditif, spline linier truncated

*(Halaman ini sengaja dikosongkan)*

# NONPARAMETRIC MIXED REGRESSION OF TRUNCATED LINEAR SPLINE AND KERNEL FUNCTION FOR POVERTY DATA MODELING IN PAPUA PROVINCE


| | |
|---|---|
| Name | : Rory |
| NRP | : 1314 201 711 |
| Supervisor | : Prof. Dr. Drs. I Nyoman Budiantara, M.Si. |
| Co-Supervisor | : Dr. Wahyu Wibowo, S.Si., M.Si. |


## ABSTRACT


Nonparametric mixed regression model $y_i = f(u_i, \tilde{v}_i) + \varepsilon_i, i = 1,2, \dots, n$, $\tilde{v}_i = (v_{1i}, v_{2i}, \dots, v_{mi})^T$ has additive regression curve $f(u_i, \tilde{v}_i) = g(u_i) + \sum_{j=1}^{m} h_j(v_{ji})$. Component of $g(u_i)$ is approached by truncated linear spline regression curve, while components of $h_j(v_{ji})$ is approached by Nadaraya-Watson kernel regression curve. Random error $\varepsilon_i$ follows independent normal distribution normal $N(0, \sigma^2)$. The aim of this thesis is to perform a study on estimator of nonparametric mixed regression curve of spline and kernel $f(u_i, \tilde{v}_i)$ and to apply it to poverty data in Papua Province. The results of study show that estimator of spline regression curve $\tilde{g}(u)$ is $\hat{\tilde{g}}_{\tilde{\phi},\tilde{\xi}}(u, \tilde{v}) = \mathbf{S}(\tilde{\xi}, \tilde{\phi})\tilde{y}$ and estimator of kernel regression curve $\sum_{j=1}^{m} \tilde{h}_j(v_j)$ is $\sum_{j=1}^{m} \hat{\tilde{h}}_j(v_j) = \mathbf{V}(\tilde{\phi})\tilde{y}$. Furthermore, estimator of nonparametric mixed regression curve of spline and kernel $\tilde{f}(u, \tilde{v})$ is $\hat{\tilde{f}}_{\tilde{\phi},\tilde{\xi}}(u, \tilde{v}) = \mathbf{Z}(\tilde{\xi}, \tilde{\phi})\tilde{y}$, where $\mathbf{Z}(\tilde{\xi}, \tilde{\phi}) = \mathbf{S}(\tilde{\xi}, \tilde{\phi}) + \mathbf{V}(\tilde{\phi})$. The estimators are biased estimator, but still the class of linear estimator. Matrix $\mathbf{S}(\tilde{\xi}, \tilde{\phi})$, $\mathbf{V}(\tilde{\phi})$ and $\mathbf{Z}(\tilde{\xi}, \tilde{\phi})$ depend on knot points location $\tilde{\xi}$ and bandwidths $\tilde{\phi}$. The best nonparametric mixed regression model of spline and kernel is the model that uses the optimum of number of knots point, knot points location and bandwidths obtained from Generalized Cross Validation (GCV) method. The choice of knot points location and bandwidths is conducted simultaneously. Nonparametric mixed regression model of spline and kernel is applied to poverty data in Papua Province, in which the response variable is percentage of poor people ($y$), the predictor variable that follows spline regression curve is GDRP per capita ($u$), and the predictor variables that follow kernel curve regression is gini ratio ($v_1$), mean years of schooling ($v_2$), unemployment rate ($v_3$) and economic growth ($v_4$). The best model is obtained when the model has 3 knot points. Estimation model results $R^2$ of 92.02%. The model can be used for policy scenario.



Key words: additive nonparametric regression, kernel nadaraya-watson, nonparametric mixed regression, truncated linear spline

*(Halaman ini sengaja dikosongkan)*