

INTISARI

ESTIMASI MODEL REGRESI NONPARAMETRIK DENGAN PENDEKATAN DERET FOURIER MENGGUNAKAN METODE OLS DAN PLS PADA DATA KEMISKINAN PROVINSI JAWA TENGAH

oleh

Tias Maharani

20/459375/PA/20036

Regresi nonparametrik adalah metode statistika untuk mengestimasi kurva regresi untuk mengetahui hubungan antara variabel prediktor terhadap variabel respon yang tidak mengikuti pola tertentu. Penelitian ini menggunakan pendekatan regresi nonparametrik dengan estimator deret Fourier karena kelebihanannya memiliki fleksibilitas dan baik untuk mendekati pola data yang berulang. Diberikan data berpasangan (y_i, x_{ji}) dengan i adalah banyaknya pengamatan dan j adalah banyaknya variabel prediktor. Model regresi nonparametrik dengan pendekatan deret Fourier yang memuat variabel tersebut dinyatakan sebagai $y_i = \frac{1}{2}a_0 + \sum_{j=1}^p [b_j x_{ji} + \sum_{k=1}^K a_{kj} \cos kx_{ji}] + \varepsilon_i$. Tujuan dari penelitian ini adalah memperoleh bentuk estimator dengan menggunakan metode *Ordinary Least Square* (OLS) dan *Penalized Least Square* (PLS). Metode OLS mengestimasi parameter regresi dengan meminimumkan jumlah kuadrat residual, hasilnya diperoleh estimator $\beta = (X^T X)^{-1} X^T y$. Metode PLS yang merupakan pengembangan dari OLS dengan penambahan penalti dan parameter penghalus yang berfungsi untuk menangani multikolinearitas dan menghindari model yang *overfitting*. Metode ini mengestimasi parameter regresi dengan meminimumkan fungsi PLS, hasilnya diperoleh estimator $\beta = (n^{-1} X^T X + \lambda D)^{-1} n^{-1} X^T y$. Parameter osilasi (K) dan parameter penghalus (λ)

ditentukan dengan nilai *Generalized Cross-Validation* (GCV) minimum, koefisien determinasi (R^2) besar, dan parameter model sederhana. Hasil estimasi selanjutnya diaplikasikan untuk memodelkan kemiskinan di Provinsi Jawa Tengah dengan variabel prediktor angka melek huruf, pendidikan kurang dari sekolah dasar (SD), dan rata-rata lama sekolah. Model kemiskinan di Provinsi Jawa Tengah yang diestimasi dengan metode OLS ditentukan nilai K sebesar 6 menghasilkan GCV sebesar 11.490476 dan R^2 sebesar 75.04464%. Sementara model yang diestimasi dengan metode PLS ditentukan nilai dengan parameter osilasi (K) sebesar 11 dengan $\lambda = 0.0000315$ yang menghasilkan GCV sebesar 4.438 dan R^2 sebesar 75.08203%.

Kata Kunci: Regresi Nonparametrik, Deret Fourier, OLS, PLS.

ABSTRACT

ESTIMATION OF NONPARAMETRIC REGRESSION MODEL WITH A FOURIER SERIES APPROACH USING OLS AND PLS METHODS ON PROVERTY DATA OF CENTRAL JAVA PROVINCE

by

Tias Maharani

20/459375/PA/20036

Nonparametric regression is a statistics method to estimate the regression curve to determine the relationship between predictor variables to responses with uncertain pattern. This research uses a nonparametric regression approach with a Fourier series estimator because of its advantages of having flexibility and being good at approximating repeated data patterns. Given paired data (y_i, x_{ji}) where i is the number of observations and j is the number of predictor variables. The nonparametric regression model with a Fourier series approach containing these variables is expressed as $y_i = \frac{1}{2}a_0 + \sum_{j=1}^p [b_j x_{ji} + \sum_{k=1}^K a_{kj} \cos kx_{ji}] + \varepsilon_i$. The aim of this research is to get an estimator using the *Ordinary Least Square* (OLS) and *Penalized Least Square* (PLS) methods. The OLS method estimates regression parameters by minimizing the sum of squared residuals, the result is the estimator $\beta = (X^T X)^{-1} X^T y$. The PLS method, which is a development of OLS with the addition of penalties and smoothing parameters to handle multicollinearity and avoid overfitting models. This method estimates regression parameters by minimizing the PLS function, the result is an estimator $\beta = (n^{-1} X^T X + \lambda D)^{-1} n^{-1} X^T y$. The oscillation parameters (K) and smoothing parameters (λ) are determined with minimum *Generalized Cross-Validation* (GCV) values, maximum coefficients of determination (R^2), and simple model

parameters. The estimation results were then applied to model poverty in Central Java Province with the predictor variables literacy rate, education less than elementary school, and average years of schooling. The poverty model in Central Java Province estimated using the OLS method determined a K value of 6 produces a GCV of 11.490476 and R^2 of 75.04464%. Meanwhile, the model estimated using the PLS method determined by a value with an oscillation parameter (K) of 11 with $\lambda = 0.0000315$, resulting in a GCV of 4.438 and an R^2 of 75.08203%.

Keywords: Nonparametric Regression, Fourier Series, OLS, PLS.