

INTISARI

Dalam disertasi ini diteliti mengenai regresi nonparametrik multivariabel spline *truncated* dalam model *geographically weighted regression* (GWR). Model tersebut merupakan pengembangan dari regresi nonparametrik spline *truncated* yang memperhitungkan faktor geografis atau spasial. Spline *Truncated* merupakan suatu fungsi yang dibangun dengan basis komponen polinomial dan komponen *truncated* yaitu potongan-potongan polinomial yang memiliki titik knot, yang dapat mengatasi pola perubahan perilaku data. Pendekatan spline *truncated* dijadikan solusi untuk menyelesaikan masalah pemodelan analisis data spasial yang hubungan antara variabel respon dan variabel prediktor tidak mengikuti pola tertentu dan terdapat pola yang berubah-ubah pada sub-sub interval tertentu. Penelitian ini diawali dengan membahas secara teoritis proses inferensi untuk mendapatkan estimator parameter dengan menggunakan metode *Maximum Likelihood Estimator* (MLE) terboboti yang berakibat diperoleh estimator fungsi regresi, kemudian menyelidiki sifat-sifat dari estimator tersebut. Penelitian selanjutnya penentuan uji hipotesis kesesuaian model antara regresi nonparametrik multivariabel spline *truncated* dalam GWR versus regresi nonparametrik multivariabel spline *truncated*. Uji hipotesis ini bertujuan untuk menentukan model apa yang paling sesuai digunakan untuk analisis data spasial. Metode yang digunakan dalam menurunkan statistik uji adalah *maximum likelihood ratio test* (MLRT). Setelah mendapatkan formulasi uji hipotesis kesesuaian model selanjutnya akan dicari statistik uji dan distribusi untuk uji hipotesis simultan untuk parameter model. Selanjutnya diterapkan aplikasi pemodelan menggunakan data tingkat pengangguran terbuka (TPT) tahun 2016 di 38 Kabupaten/Kota di Jawa Timur.

Hasil penelitian diperoleh fungsi likelihood yang dibangun berbeda untuk setiap lokasi pengamatan, perbedaan diberikan oleh pembobot yang bergantung pada jarak antar titik lokasi pengamatan. Akibatnya estimator parameter pada setiap lokasi pengamatan berbeda dengan estimator parameter pada lokasi pengamatan yang lain. Estimator tersebut memuat komponen polinomial yang direpresentasikan oleh matriks $\mathbf{X}$ dan komponen *truncated* yang direpresentasikan oleh matriks $\mathbf{P}$ dan matriks pembobot $\mathbf{W}(u_i, v_i)$ yang merupakan matriks diagonal yang setiap elemennya adalah jarak yang bernilai positif sehingga $\mathbf{W}(u_i, v_i)$ didefinisikan sebagai matriks definit positif. Karena elemen-elemen kolom matriks $(\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1}$ dan $(\mathbf{P}^T \mathbf{W}(u_i, v_i) \mathbf{P})^{-1}$ independent linear dan merupakan matriks dengan rank penuh (full rank) maka sifat ini yang mengakibatkan estimator parameter mempunyai invers. Hasil selanjutnya sebagai akibat dari ditemukan estimator parameter diperoleh estimator kurva regresi $\hat{f}$. Sifat-sifat estimator $\hat{\beta}(u_i, v_i)$, $\hat{\delta}(u_i, v_i)$ dan $\hat{f}$ masing-masing menunjukkan estimator tak bias dan tergolong estimator linier dalam observasi $\tilde{\mathbf{Y}}$.

Uji hipotesis kesesuaian model dihasilkan statistik uji yang dibangun berdasarkan rasio likelihood dari himpunan parameter di bawah Hipotesis nol yang komponennya terdiri atas parameter yang tidak dipengaruhi oleh faktor geografis dan himpunan di bawah parameter populasi yang komponennya terdiri dari parameter yang dipengaruhi oleh faktor geografis. Selanjutnya untuk membuktikan distribusi dari statistik uji, terlebih dahulu kita akan membuktikan masing-masing pembilang dan penyebut dalam statistik uji berdistribusi chi-kuadrat. Permasalahan terdapat pada matriks $D(u_i, v_i)$ yang bukan matriks idempotent, karena memuat matriks pembobot $W(u_i, v_i)$ yang nilainya berbeda-beda setiap lokasi pengamatan. Permasalahan tersebut dapat diselesaikan dengan modifikasi matematis dan menghasilkan suatu konstruksi matematis yang baru sehingga dihasilkan statistik uji berdistribusi F . Bentuk statistik uji untuk uji simultan berdistribusi F dan uji parsial berdistribusi t .

Hasil studi empirik menunjukkan bahwa data TPT mempunyai pengaruh geografis dan terdapat pola yang berubah-ubah pada sub-sub interval tertentu. Berdasarkan hasil uji hipotesis kesesuaian model, model yang sesuai digunakan adalah regresi nonparametrik multivariabel spline *truncated* dalam GWR. Pembobot yang digunakan adalah fungsi Kernel Gaussian dengan nilai *Cross validation* (CV) lebih rendah dibandingkan fungsi Kernel Bisquare. Nilai GCV terendah diperoleh pada pemodelan dengan satu titik knot, pemodelan tersebut menghasilkan koefisien determinasi sebesar 80,42%. Berdasarkan hasil statistik uji simultan diperoleh kesimpulan paling sedikit terdapat satu parameter pada model regresi nonparametrik multivariable spline *truncated* dalam model GWR yang signifikan terhadap variabel respon. Selanjutnya Statistik Uji parsial menghasilkan 10 pengelompokan area berdasarkan variabel predictor yang signifikan pada masing-masing lokasi penelitian.

Kata Kunci: Estimator, Fungsi kernel Gaussian, Fungsi kernel bisquare, GWR, MLE, MLRT, Pembobot, Regresi nonparametrik, Spline *truncated*, Statistik uji, TPT, Titik knot, Uji hipotesis.

ABSTRACT

This dissertation examines the multivariable nonparametric truncated spline regression in Geographically Weighted Regression (GWR) model. The model is an expansion of nonparametric truncated spline regression that takes into account geographical or spatial factors. Truncated spline is a function constructed on the basis of polynomial components and truncated components i.e polynomial pieces that have knot points, which can overcome the pattern of changes in data behavior. The truncated spline approach is used as a solution to solve spatial data modeling problems where the relationship between response variables and predictor variables does not follow a particular pattern and there are patterns that vary across sub-intervals. This study began by discussing theoretically the inference process to obtain parameter estimator using The Weighted Maximum Likelihood Estimator (MLE) method which resulted in estimator of regression function, then the properties of the estimator were investigated. Subsequently, a hypothesis test of model conformity was established between multivariable nonparametric truncated spline regression in GWR versus multivariable nonparametric truncated spline regression. This hypothesis test aims to determine what model is the most suitable for spatial data analysis. The method used in deriving the test statistic was the maximum likelihood ratio test (MLRT). After obtaining the formulation of the model conformity hypothesis test, determined test and distribution statistics for hypothesis testing simultaneously for model parameters. Finally, the modeling was applied using Open Unemployment Rate (OUR) data in 2016 in 38 districts/cities in East Java.

The result of the study is the likelihood function constructed is different for each location of observation, the difference was given by the weights that depend on the distance between points of observation location. As a result, the parameter estimator at each observation location is different from the parameter estimator at the other observation location. The estimation contains a polynomial component represented by matrix $\mathbf{X}$, the truncated component is represented by matrix $\mathbf{P}$, and weighted matrix $\mathbf{W}(u_i, v_i)$ which is a diagonal matrix has positive values, therefore $\mathbf{W}(u_i, v_i)$ is defined as a matrix definite positive. Since the elements of the matrix columns $(\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1}$ and $(\mathbf{P}^T \mathbf{W}(u_i, v_i) \mathbf{P})^{-1}$ are linear independent and are full-rank matrices then this property causes the parameter estimator to have an inverse. As a result of finding parameter estimator, it was obtained the estimator of regression curve $\hat{f}$. The properties of the estimators $\hat{\beta}(u_i, v_i)$, $\hat{\delta}(u_i, v_i)$ and $\hat{f}$ respectively show unbiased estimators and are classified as linear estimators in observation $\tilde{\mathbf{Y}}$.

The model conformity hypothesis test yields test statistic constructed on likelihood ratio of parameter set under null hypothesis whose components consist of parameters not influenced by geographical factor and set under population

parameter whose components consist of parameters influenced by geographical factor. To prove the distribution of test statistic, we first proved the numerator and denominator respectively in test statistic are Chi-Square distributed. Problems are found in matrix $D(u_i, v_i)$ which is not an idempotent matrix, because it contains a weighting matrix of $W(u_i, v_i)$ whose values different from each observation location. These problems can be solved by mathematical modification and produce a new mathematical construction so that the statistics test are distributed with Distribution F . The form of test statistics for simultaneous test is Distributed F and partial test is Distributed t .

The results of the empirical study indicates that the OUR data has a geographical influence and there are patterns that vary in sub-intervals. Based on the hypothesis test of model conformity, the appropriate model used is multivariable nonparametric truncated spline regression in GWR. The weight used was the Gaussian Kernel function with the value of Cross Validation (CV) was lower than the CV of Kernel Bisquare function. The lowest GCV value was obtained on modeling with one knot point, the modeling resulted in a determination coefficient of 80.42%. Based on simultaneous statistical test results, it could be concluded that there was at least one significant parameter in the multivariable nonparametric truncated spline regression in GWR model to the response variable. Partial test statistics resulted in 10 grouping areas based on significant predictor variables at each observation location.

Keywords: Estimator, Gaussian Kernel Function, Bisquare Kernel Function, GWR, MLE, MLRT, Weighting, Nonparametric Regression, Truncated Spline, Test Statistic, OUR, Knot Point, Hypothesis Test.