

***GEOGRAPHICALLY WEIGHTED REGRESSION PRINCIPAL COMPONENT
ANALYSIS (GWRPCA) PADA PEMODELAN PENDAPATAN ASLI DAERAH
DI JAWA TENGAH***



SKRIPSI

Disusun Oleh :

NURMALITA SARI

240102120008

**DEPARTEMEN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG**

2016

***GEOGRAPHICALLY WEIGHTED REGRESSION PRINCIPAL COMPONENT
ANALYSIS (GWRPCA) PADA PEMODELAN PENDAPATAN ASLI DAERAH
DI JAWA TENGAH***

Oleh

Nurmalita Sari

24010212120008

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Sains pada Departemen Statistika**

**DEPARTEMEN STATISTIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
SEMARANG**

2016

HALAMAN PENGESAHAN I

Judul : *Geographically Weighted Regression Principal Component Analysis (GWRPCA) pada Pemodelan Pendapatan Asli Daerah di Jawa Tengah*

Nama : Nurmalita Sari

NIM : 24010212120008

Departemen : Statistika

Telah diujikan pada sidang Tugas Akhir dan dinyatakan lulus pada tanggal 10 Agustus 2016.

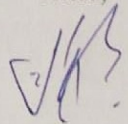
Semarang, Agustus 2016

Mengetahui,
Ketua Departemen Statistika
PSM dan DIP



Dra. Dwi Ispriyanti, M.Si
NIP. 195709141986032001

Panitia Penguji Ujian Tugas Akhir
Ketua,



Dra. Suparti, M.Si
NIP. 196509131990032001

HALAMAN PENGESAHAN II

Judul : *Geographically Weighted Regression Principal Component Analysis (GWRPCA)* pada Pemodelan Pendapatan Asli Daerah di Jawa Tengah

Nama : Nurmalita Sari

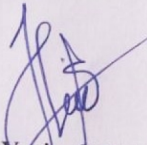
NIM : 24010212120008

Departemen : Statistika

Telah diujikan pada sidang Tugas Akhir tanggal 10 Agustus 2016.

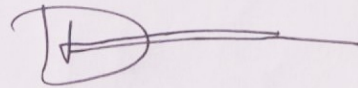
Semarang, Agustus 2016

Pembimbing I



Hasbi Yasin, S.Si, M.Si
NIP. 198212172006041003

Pembimbing II



Alan Prahutama, S.Si, M.Si
NIP. 1988042120140401002

KATA PENGANTAR

Puji syukur penulis ucapkan atas kehadiran Allah SWT yang telah melimpahkan kasih dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir yang berjudul ***“Geographically Weighted Regression Principal Component Analysis (GWRPCA) pada Pemodelan Pendapatan Asli Daerah di Jawa Tengah”***. Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Ibu Dra. Dwi Ispriyanti, M.Si selaku Ketua Departemen Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
2. Bapak Hasbi Yasin, S.Si, M.Si sebagai pembimbing I dan Bapak Alan Prahutama, S.Si, M.Si sebagai pembimbing II yang telah memberikan bimbingan dan pengarahan dalam penulisan laporan ini.
3. Bapak dan Ibu dosen Departemen Statistika Fakultas Sains dan Matematika Universitas Diponegoro yang telah memberikan ilmu yang bermanfaat.
4. Semua pihak yang tidak dapat disebutkan satu per satu yang telah mendukung penulis menyelesaikan penulisan laporan ini.

Penulis menyadari bahwa penulisan tugas akhir ini masih jauh dari sempurna.

Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan.

Semarang, Agustus 2016

Penulis

ABSTRAK

Analisis Regresi Linier Berganda merupakan metode untuk memodelkan hubungan antara sebuah variabel respon dengan dua atau lebih variabel bebas. *Geographically Weighted Regression* (GWR) merupakan pengembangan dari model regresi dimana setiap lokasi pengamatan mempunyai nilai parameter regresi yang berbeda-beda akibat adanya efek heterogenitas spasial. *Regression Principal Component Analysis* (RPCA) merupakan gabungan dari metode PCA dan regresi untuk menghilangkan efek multikolinieritas pada model regresi. *Geographically Weighted Regression Principal Component Analysis* (GWRPCA) merupakan gabungan dari metode PCA dan GWR apabila terjadi heterogenitas spasial dan multikolinieritas lokal. Estimasi parameter untuk model GWR dan GWRPCA menggunakan metode *Weighted Least Square* (WLS). Pembobotan menggunakan fungsi kernel *fixed gaussian* melalui pemilihan *bandwidth* yang optimum sebesar 0,08321242 dengan kriteria *Cross Validation* (CV) minimum sebesar 3,009035. Dalam PCA didapatkan variabel-variabel yang mempengaruhi Pendapatan Asli Daerah (PAD) di Jawa Tengah tahun 2012 dan 2013 dapat diwakili oleh PC₁ yang mampu menerangkan total varian data 71,4%. GWRPCA merupakan model terbaik untuk memodelkan PAD di kabupaten dan kota di Jawa Tengah dibandingkan model RPCA karena memiliki nilai *Akaike Information Criterion* (AIC) terkecil dan R² terbesar.

Kata Kunci: Heterogenitas Spasial, Multikolinieritas Lokal, *Principal Component Analysis*, *Geographically Weighted Regression*, *Principal Component Analysis*.

ABSTRACT

Linear Regression Analysis is a method for modeling the relation between a response variable with two or more independent variables. Geographically Weighted Regression (GWR) is a development of the regression model where each observation location has different regression parameter values because of the effects of spatial heterogeneity. Regression Principal Component Analysis (PCA) is a combination of PCA and are used to remove the effect of multicollinearity in regression. Geographically Weighted Regression Principal Component Analysis (GWRPCA) is a combination of PCA and GWR if spatial heterogeneity and local multicollinearity occurred. Estimation parameters for the GWR and GWRPCA using Weighted Least Square (WLS). Weighting use fixed gaussian kernel function through selection of the optimum bandwidth is 0,08321242 with minimum Cross Validation (CV) is 3,009035. There are some variables in PCA that affect locally-generated revenue in Central Java on 2012 and 2013, which can be represented by PC_1 that explained the total variance data about 71,4%. GWRPCA is a better model for modeling locally-generated revenue for the districts and cities in Central Java than RPCA because it has the the smallest Akaike Information Criterion (AIC) and the largest R^2 :

Keywords : Spatial Heterogeneity, Local Multicollinearity, Principal Component Analysis, Geographically Weighted Regression Principal Component Analysis.

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
KATA PENGANTAR	iv
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xiii
 BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	6
 BAB II TINJAUAN PUSTAKA	
2.1 Analisis Regresi Linier Berganda	7
2.1.1 Estimasi Parameter Model Regresi Linier Berganda ..	8
2.1.2 Uji Hipotesis dalam Regresi Linier Berganda	10
2.1.3 Uji Asumsi Analisis Regresi Linier Berganda.....	11
2.2 Uji Asumsi <i>Principal Component Analysis</i> (PCA).....	14

2.2.1	<i>Kaiser-Mayer-Olkin (KMO)</i>	15
2.2.2	<i>Measure Sampling of Adequacy (MSA)</i>	16
2.3	<i>Principal Component Analysis (PCA)</i>	17
2.4	<i>Geographically Weighted Regression (GWR)</i>	19
2.4.1	Estimasi Parameter Model GWR.....	20
2.4.2	Koordinat Spasial	22
2.4.3	Pemilihan <i>Bandwidth</i> dan Pembobot	22
2.5	<i>Geographically Weighted Regression Principal Component Analysis (GWRPCA)</i>	24
2.5.1	Pengujian Kesesuaian Model GWRPCA	26
2.5.2	Pengujian Parsial Parameter Model GWRPCA	27
2.6	Pemilihan Model Terbaik	29
2.7	Pendapatan Asli Daerah (PAD)	31
2.8	Faktor-Faktor yang Mempengaruhi PAD	31
2.8.1	Jumlah Penduduk	31
2.8.2	Retribusi Daerah	32
2.8.3	Belanja Daerah	32
2.8.4	PDRB	33
2.8.5	Pajak Daerah	34

BAB III METODOLOGI PENELITIAN

3.1	Sumber Data	36
3.2	Variabel Penelitian	36
3.3	Langkah Penelitian	37

3.4	Diagram Alir Analisis (<i>Flowchart</i>)	39
BAB IV ANALISIS DAN PEMBAHASAN		
4.1	Deskripsi Data	41
4.2	Analisis Regresi Linier Berganda	43
4.2.1	Uji Hipotesis dalam Regresi Linier Berganda	44
4.2.2	Uji Normalitas <i>Error</i>	45
4.2.3	Uji Heterogenitas Spasial.....	46
4.2.4	Uji Multikolinieritas.....	47
4.3	Pemodelan Menggunakan GWR.....	48
4.4	Pengujian Multikolinieritas Lokal pada Model GWR	49
4.5	Uji Asumsi <i>Principal Component Analysis</i>	50
4.5.1	<i>Kaiser-Mayer-Olkin</i> (KMO)	51
4.5.2	<i>Measure Sampling of Adequacy</i> (MSA)	51
4.6	Pembentukan Komponen Utama	52
4.7	Pemodelan Menggunakan RPCA	54
4.7.1	Uji Signifikansi RPCA	55
4.7.2	Pengujian Koefisien RPCA secara Individu	55
4.8	Pemodelan Menggunakan GWRRPCA	56
4.8.1	Pengujian Kesesuaian Model GWPCA.....	58
4.8.2	Pengujian Parsial Parameter Model GWRPCA	59
4.9	Pemilihan Model Terbaik.....	60
BAB V PENUTUP		63

DAFTAR PUSTAKA	64
LAMPIRAN	67

DAFTAR GAMBAR

	Halaman
Gambar 1. Diagram Alir Analisis Data Pengamatan	40
Gambar 2. Grafik PAD di Jawa Tengah Tahun 2012 dan 2013	42

DAFTAR TABEL

	Halaman
Tabel 1. Kriteria Nilai MSA	16
Tabel 2. Variabel Penelitian	36
Tabel 3. Deskriptif Data PAD di Jawa Tengah Tahun 2012 dan 2013	43
Tabel 4. Uji Koefisien Regresi	45
Tabel 5. Nilai VIF.....	47
Tabel 6. Pemilihan Fungsi Kernel Berdasarkan AIC dan R^2	48
Tabel 7. Nilai MSA	52
Tabel 8. Pemilihan Model Terbaik.....	61

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data PAD di Jawa Jawa Tengah tahun 2012	67
Lampiran 2. Data PAD di Jawa Jawa Tengah tahun 2013	68
Lampiran 3. Data PAD di Jawa Jawa Tengah tahun 2012 yang Distandarkan	69
Lampiran 4. Data PAD di Jawa Jawa Tengah tahun 2013 yang Distandarkan	70
Lampiran 5. Letak Titik Koordinat Kabupaten dan Kota di Jawa Tengah	71
Lampiran 6. Output Analisis Regresi Linier Berganda	72
Lampiran 7. Uji Normalitas <i>Error</i>	73
Lampiran 8. Uji Heterogenitas Spasial	74
Lampiran 9. Jarak <i>Euclidean</i> pada Setiap Lokasi Kabupaten dan Kota	75
Lampiran 10. Contoh Komponen Diagonal Matriks Pembobot Fungsi Kernel Fixed Gaussian pada Kota Semarang dengan Menggunakan <i>Bandwidth</i> sebesar 0,2054175 pada pemodelan GWR	78
Lampiran 11. Estimasi Parameter Model GWR pada PAD di Jawa Tengah...	79
Lampiran 12. Nilai VIF pada Variabel Bebas Model GWR.....	80
Lampiran 13. Uji Kecukupan Data	81
Lampiran 14. Output <i>Principal Component Analysis</i> (PCA).....	82
Lampiran 15. Output RPCA.....	83
Lampiran 16. Data PAD Menggunakan PCA	84
Lampiran 17. Uji Heterogenitas Spasial menggunakan variabel PC_1	86
Lampiran 18. Output Program GWRPCA untuk PAD di Jawa Tengah	87

Lampiran 19. Contoh Komponen Diagonal Matriks Pembobot Fungsi Kernel Fixed Gaussian pada Kabupaten Cilacap dengan Menggunakan <i>Bandwidth</i> sebesar 0,08321242 pada pemodelan GWRPCA	89
Lampiran 20. Estimasi Parameter Model GWRPCA PAD di Jawa Tengah.... .	90
Lampiran 21. Model GWRPCA PAD di Jawa Tengah..... .	92
Lampiran 22. Model GWRPCA PAD di Jawa Tengah yang Distandarkan	93
Lampiran 23. Output Pengujian Hipotesis Model GWRPCA	94
Lampiran 24. Koefisien Determinasi Lokal pada Pemodelan PAD di Jawa Tengah Menggunakan GWRPCA	95

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan suatu metode statistika yang menganalisis hubungan antara dua variabel atau lebih yang umumnya dinyatakan dalam persamaan matematik. Analisis regresi digunakan untuk mengetahui pengaruh dua atau lebih variabel bebas (independen) terhadap sebuah variabel respon (dependen). Menurut Ifadah (2011) dalam statistika sebuah model regresi dikatakan baik atau cocok, jika memenuhi asumsi-asumsi ideal (klasik), antara lain tidak adanya autokorelasi, heteroskedastisitas dan multikolinieritas. Sehingga proses kontrol terhadap model perlu dilakukan untuk menelaah dipenuhi tidaknya asumsi tersebut. Dengan terpenuhinya asumsi tersebut, maka hasil yang diperoleh dapat lebih akurat dan mendekati atau sama dengan kenyataan.

Masalah yang sering kali muncul dalam analisis regresi yaitu terdapatnya multikolinieritas di antara variabel bebas. Multikolinieritas terjadi ketika di dalam model regresi ada beberapa atau semua variabel bebas mempunyai hubungan linier sempurna atau hampir sempurna. Metode alternatif yang dapat digunakan untuk mengatasi permasalahan multikolinieritas dalam analisis regresi yaitu dengan metode Analisis Komponen Utama atau *Principal Component Analysis* (PCA). Tujuan penggunaan metode PCA adalah untuk menyederhanakan (mereduksi) variabel-variabel bebas yang mempengaruhi variabel respon dengan

mempertahankan sebanyak mungkin informasi dan menghilangkan korelasi antara variabel-variabel tersebut.

Beberapa penelitian telah dilakukan mengenai penggunaan *Regression Principal Component Analysis* (RPCA), salah satunya oleh Tazliqoh *et al.* (2015) yang menggunakan perbandingan metode Regresi Komponen Utama dengan Regresi Ridge untuk menganalisis faktor-faktor yang mempengaruhi Pendapatan Asli Daerah (PAD) di Jawa Tengah. Dalam penelitian tersebut diperoleh hasil bahwa jumlah penduduk, retribusi daerah, belanja daerah, PDRB atas harga konstan, PDRB atas harga berlaku, dan pajak mempengaruhi PAD di Jawa Tengah. Selain itu, hasil analisis terhadap faktor-faktor yang mempengaruhi PAD menunjukkan bahwa terdapat masalah multikolinieritas sehingga perlu penanganan menggunakan metode regresi komponen utama dan regresi ridge. Perbandingan antara kedua metode menghasilkan bahwa Regresi Komponen Utama (*Regression Principal Component Analysis*) lebih baik digunakan untuk menganalisis faktor-faktor yang mempengaruhi PAD di Jawa Tengah dibandingkan metode regresi ridge.

Selain itu, pada model regresi diasumsikan bahwa nilai parameter regresi akan selalu tetap (konstan). Sehingga parameter regresi disebut juga dengan parameter global. Artinya parameter regresi akan mempunyai nilai yang sama untuk setiap wilayah pengamatan. Apabila diduga ada pengaruh unsur wilayah dalam model regresi, maka metode regresi akan kurang tepat dilakukan.

Adanya unsur wilayah atau biasa disebut spasial tidak dapat dianalisis secara global karena setiap wilayah mempunyai karakteristik sendiri yang

mungkin berbeda antara wilayah yang satu dengan yang lainnya. Perbedaan karakteristik wilayah yang satu dengan wilayah yang lainnya dapat diketahui dengan dilakukan uji heterogenitas spasial. Heterogenitas muncul jika *error* atau residual dari model yang diamati memiliki varian yang tidak konstan dari satu observasi ke observasi lainnya.

Menurut Isbiyantoro *et al.* (2014) pada data amatan yang berupa data spasial, kenyataannya kondisi semua wilayah yang diamati tidak sama karena adanya faktor geografis, keadaan sosial budaya, maupun hal lainnya yang melatarbelakangi kondisi yang diteliti. Bila kasus ini terjadi, maka regresi linier biasa kurang mampu dalam menjelaskan fenomena data yang sebenarnya. Salah satu metode untuk mengatasi masalah heterogenitas spasial yaitu dengan menggunakan metode *Geographically Weighted Regression* (GWR).

GWR merupakan pengembangan dari model regresi dimana setiap parameter dihitung berdasarkan setiap lokasi pengamatan, sehingga setiap lokasi pengamatan mempunyai nilai parameter regresi yang berbeda-beda atau bersifat lokal (Fotheringham *et al.*, 2002). Untuk memberikan hasil penaksiran parameter yang berbeda pada lokasi berbeda dibutuhkan fungsi pembobot dengan *bandwidth* optimum yang sesuai dengan kondisi data. Peran pembobot dalam GWR menjadi penting, karena nilai pembobot mewakili letak data observasi antara satu dengan lainnya.

Pendapatan Asli Daerah (PAD) biasanya ditunjukkan dengan kecenderungan wilayah yang mempunyai kekayaan dan potensi-potensi yang melimpah, sehingga perlu ditambahkan informasi geografis dari lokasi daerah

yang diamati. Hal ini melibatkan efek spasial dalam menganalisis faktor-faktor yang mempengaruhi PAD. Pada data spasial kondisi lokasi yang satu dengan lokasi lain yang tidak sama, yang ditinjau dari segi geografis, keadaan sosial-budaya maupun hal lain yang dapat menimbulkan kondisi heterogenitas spasial pada lokasi yang diteliti. Banyaknya faktor yang mempengaruhi PAD memungkinkan terdapatnya multikolinieritas lokal, yaitu korelasi antara variabel bebas yang satu dengan variabel bebas lainnya pada masing-masing wilayah amatan. Sehingga kemungkinan pada data Pendapatan Asli Daerah (PAD) terdapat masalah heterogenitas spasial dan multikolinieritas lokal.

Salah satu metode untuk mengatasi masalah heterogenitas spasial dan multikolinieritas lokal pada pemodelan Pendapatan Asli Daerah di Jawa Tengah yaitu dengan menggunakan *Geographically Weighted Regression Principal Component Analysis* (GWRPCA). Hasil dari analisis GWRPCA adalah pemodelan komponen utama pada variabel-variabel bebas yang mempengaruhi PAD pada masing-masing lokasi dengan menggunakan pembobot dan *bandwidth* optimum.

PAD memiliki peran yang cukup penting dalam menentukan kemampuan daerah untuk melakukan aktivitas pemerintahan dan program-program pembangunan bagi daerahnya masing-masing. Pemerintah daerah diharapkan mampu meningkatkan PAD. Salah satunya dengan mengoptimalkan faktor-faktor yang mempengaruhi PAD. Berdasarkan uraian tersebut, peneliti ingin membandingkan *Regression Principal Component Analysis* (RPCA) dan *Geographically Weighted Regression Principal Component Analysis* (GWRPCA)

mengenai faktor-faktor yang mempengaruhi Pendapatan Asli Daerah (PAD) di Jawa Tengah antara lain jumlah penduduk, retribusi daerah, belanja daerah, PDRB, dan pajak daerah.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana pemodelan Pendapatan Asli Daerah (PAD) di Jawa Tengah menggunakan metode *Geographically Weighted Regression* (GWR)?
2. Bagaimana membentuk variabel-variabel baru (komponen utama) yang tidak berkorelasi sehingga dapat mewakili atau menggantikan variabel-variabel yang berpengaruh terhadap Pendapatan Asli Daerah (PAD) di Jawa Tengah menggunakan metode *Regression Principal Component Analysis* (RPCA)?
3. Bagaimana pemodelan Pendapatan Asli Daerah (PAD) di Jawa Tengah menggunakan metode *Geographically Weighted Regression Principal Component Analysis* (GWRPCA)?
4. Berdasarkan studi kasus yang diteliti, metode manakah antara RPCA dan GWRPCA yang lebih efektif?

1.3 Batasan Masalah

Pada penelitian ini, masalah akan dibatasi mengenai penyederhanaan (reduksi) enam variabel yang diduga mempengaruhi Pendapatan Asli Daerah (PAD) di Jawa Tengah pada tahun 2012 dan 2013 serta pembentukan model

menggunakan metode RPCA dan GWRPCA dengan matriks pembobot *gaussian*, *bisquare*, *exponential*, dan *tricube*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menentukan model Pendapatan Asli Daerah (PAD) di Jawa Tengah menggunakan metode *Geographically Weighted Regression* (GWR).
2. Mereduksi banyaknya variabel-variabel yang berpengaruh terhadap Pendapatan Asli Daerah (PAD) di Jawa Tengah yang saling berkorelasi menjadi sedikit variabel baru (komponen utama) dengan metode *Regression Principal Components Analysis* (RPCA).
3. Menentukan model Pendapatan Asli Daerah (PAD) di Jawa Tengah dengan memperhatikan variasi spasial menggunakan metode *Geographically Weighted Regression Principal Components Analysis* (GWRPCA).
4. Mengetahui metode yang efektif untuk mengatasi masalah permasalahan heterogenitas spasial dan multikolinieritas lokal dengan menggunakan metode RPCA dan GWRPCA.