



## PERBANDINGAN METODE *PARTIAL LEAST SQUARE* (PLS) DAN *PRINCIPAL COMPONENT REGRESSION* (PCR) UNTUK MENGATASI MULTIKOLINEARITAS PADA MODEL REGRESI LINEAR BERGANDA

Eko Supriyadi<sup>✉</sup>, Scolastika Mariani, Sugiman

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia  
Gedung D7 Lt. 1, Kampus Sekaran Gunungpati, Semarang 50229

### Info Artikel

Sejarah Artikel:  
Diterima Juli 2017  
Disetujui Agustus 2017  
Dipublikasikan Nopember 2017

#### Keywords:

Multikolinearitas;  
*Partial Least Square* (PLS);  
*Principal Component Regression* (PCR)

### Abstrak

Salah satu asumsi analisis regresi linear berganda yaitu tidak terjadi masalah multikolinearitas. Apabila terjadi masalah multikolinearitas, metode *Partial Least Square* (PLS) dan *Principal Component Regression* (PCR) merupakan dua metode yang dapat digunakan untuk mengatasi masalah multikolinearitas tersebut. Metode yang digunakan dalam penelitian ini adalah *Partial Least Square* (PLS) dan *Principal Component Regression* (PCR) dengan data anggaran pendapatan Daerah Provinsi Jawa Tengah 2013. Hasil dari penelitian ini diperoleh model persamaan regresi dengan metode *Partial Least Square* (PLS) dan model persamaan regresi dengan metode *Principal Component Regression* (PCR); Selanjutnya dipilih metode terbaik dengan menggunakan kriteria nilai  $R^2$  tertinggi dan MSE terkecil. Pemilihan metode terbaik adalah dengan melihat nilai  $R^2$  tertinggi dan MSE terkecil. Dapat disimpulkan bahwa metode yang lebih baik adalah *Partial Least Square* dengan nilai  $R^2 = 0,7752$  dan nilai MSE yang dihasilkan *Partial Least Square* = 3,660671E16.

### Abstract

*One of analysis assumptions of bifilar linier regression is there is no multikolinearity problem occurs. If there is a multikolinearity problem occurs, Partial Least Square (PLS) and Principal Component Regression (PCR) methods are can used to solve the multikolinearity problem. The method that was used in this research were Partial Least Square (PLS) and Principal Component Regression (PCR) by using the data of budgeting income of Central Java in 2013. The result of this research was gotten a model of regression equation by using Partial Least Square (PLS) method and a model of regression equation by using Principal Component Regression (PCR); After that choose the best method selection criteria was by seeing at the highest  $R^2$  score and the lowest MSE. The best method selection was by seeing at the highest  $R^2$  score and the lowest MSE. Thus, it could be concluded that Partial Least Square with the score of  $R^2 = 0,7752$  and the MSE score that was resulted by Partial Least Square = 3,660671E16.*

### How to Cite

Supriyadi, E., Mariani, S. & Sugiman. (2017). Perbandingan Metode Partial Least Square (PLS) dan Principal Component Regression (PCR) untuk Mengatasi Multikolinearitas pada Model Regresi Linear Berganda. *Unnes Journal of Mathematics*, 6(2): 117-128.

## PENDAHULUAN

Analisis regresi merupakan analisis yang mempelajari bagaimana membentuk sebuah hubungan fungsional dari data untuk dapat menjelaskan atau meramalkan suatu fenomena alami atas dasar fenomena yang lain. Analisis regresi memiliki peranan yang penting dalam berbagai bidang ilmu pengetahuan. Kebanyakan analisis regresi bergantung pada metode kuadrat terkecil untuk mengestimasi parameter-parameternya dalam model regresi. Tetapi metode ini biasanya dibentuk dengan beberapa asumsi, seperti tidak ada autokorelasi, tidak terjadi multikolinearitas, homoskedastisitas, dan *error* berdistribusi normal.

Analisis regresi linear berganda yang mempunyai banyak variabel bebas sering timbul masalah karena terjadinya hubungan antara dua atau lebih variabel bebasnya. Variabel bebas yang saling berkorelasi disebut multikolinearitas (*multicollinearity*). Salah satu dari asumsi model regresi linear adalah bahwa tidak terdapat multikolinearitas diantara variabel bebas yang termasuk dalam model. Multikolinearitas terjadi apabila terdapat hubungan atau korelasi diantara beberapa atau seluruh variabel bebas (Gonst and Mason, 1977 dalam Soemartini, 2008).

Pada pembentukan model regresi terdapat kemungkinan adanya hubungan antara variabel satu dengan variabel yang lain. Multikolinearitas menyebabkan estimator mempunyai varian yang besar akibatnya interval estimasi cenderung lebih besar sehingga membuat variabel bebas secara statistika tidak signifikan padahal nilai koefisien determinasi ( $R^2$ ) tinggi sehingga sulit mendapatkan estimasi yang tepat (Widarjono, 2007). Oleh karena itu perlu dilakukan tindakan lanjut untuk menangani multikolinearitas.

Efek multikolinearitas dapat menjadikan nilai model tidak dapat menjelaskan hubungan antara variabel *dependent* dan variabel *independent* secara baik. Keberadaan multikolinearitas akan menyebabkan varians parameter yang diestimasi akan menjadi lebih besar dari yang seharusnya dengan demikian tingkat akurasi dari estimasi akan menurun (Sukmono, 2014).

Multikolinearitas dalam suatu model regresi dapat diketahui dengan menghitung nilai *Variance Inflation Factor* (VIF). VIF adalah suatu faktor yang mengukur seberapa besar kenaikan ragam dari koefisien penduga regresi dibandingkan terhadap variabel bebas yang orthogonal jika dihubungkan secara linear. Nilai VIF akan semakin besar jika terdapat korelasi yang semakin besar diantara variabel bebas. Nilai VIF > 10 dapat digunakan sebagai petunjuk adanya multikolinearitas. Gejala

multikolinearitas menimbulkan masalah dalam model regresi. Korelasi antar variabel bebas yang sangat tinggi menghasilkan penduga model regresi yang bias, tidak stabil, dan mungkin jauh dari nilai prediksinya (Farahani *et al.*, 2010).

Multikolinearitas dapat diatasi dengan beberapa metode antara lain *Partial Least Square* (PLS) dan *Principal Component Regression* (PCR) (Saika, 2014). Metode PLS merupakan metode yang mengkombinasikan sifat-sifat dari komponen utama dan regresi linear berganda. Tujuan dari metode PLS adalah mengestimasi dan menganalisis variabel terikat dari variabel-variabel bebas. Dalam hal ini, PLS mereduksi dimensi variabel-variabel bebas dengan membentuk variabel-variabel baru yang merupakan kombinasi linear dari variabel-variabel bebas dengan dimensi lebih kecil (Abdi, 2010). Sedangkan metode PCR merupakan salah satu analisis regresi yang menggunakan komponen utama untuk mengatasi masalah multikolinieritas pada regresi berganda (Tazliqoh *et al.*, 2015). Komponen utama merupakan suatu teknik statistika untuk mengubah dari sebagian besar variabel asli yang digunakan yang saling berkorelasi satu dengan yang lainnya menjadi satu kumpulan variabel baru yang lebih kecil dan saling bebas (Johnson & Wichern, 2010).

Menurut Nurhasanah (2012) hasil penelitian menunjukkan metode *Partial Least Square* (PLS) nilai koefisien penduga pada masing-masing variabel tidak semuanya berpengaruh nyata pada taraf nyata 0,05 sedangkan pada regresi komponen utama semua nilai koefisien penduga pada masing-masing variabel semuanya berpengaruh nyata pada taraf nyata 0,05. Metode *Partial Least Square* (PLS) memberikan hasil yang lebih baik jika dibandingkan dengan metode regresi komponen utama. Hal ini dapat disimpulkan dengan melihat nilai ( $R^2$ ). *Mean Square Error Prediction* (MSEP), dan *Root Mean Square Error Prediction* (RMSEP). Metode *Partial Least Square* (PLS) mempunyai nilai ( $R^2$ ) yang lebih tinggi dan mempunyai nilai MSEP dan RMSEP yang lebih rendah jika dibandingkan terhadap metode regresi komponen utama.

Dari uraian di atas maka penulis menganalisis dengan judul "Perbandingan metode *Partial Least Square* (PLS) dan *Principal Component Regression* (PCR) untuk mengatasi multikolinearitas pada model regresi linear berganda". Dari kedua metode *Partial Least Square* (PLS) dan *Principal Component Regression* (PCR) akan membandingkan nilai koefisien determinasi ( $R^2$ ) dan *Mean Square Error* (MSE).

Kemudian didapatkan metode mana yang lebih baik untuk mengatasi multikolinearitas.

Penelitian ini didukung dengan penggunaan paket program SAS. Paket program SAS (*Statistical Analysis System*) adalah paket program yang mendukung analisis dalam bidang statistika, riset operasi, analisis ekonomi, time series dan lain-lain.

Rumusan masalah dalam penelitian ini adalah (1) Bagaimana model persamaan regresi dengan metode *Partial Least Square* (PLS) untuk mengatasi multikolinearitas? (2) Bagaimana model persamaan regresi dengan metode *Principal Component Regression* (PCR) untuk mengatasi multikolinearitas? (3) Metode manakah antara *Partial Least Square* (PLS) dan *Principal Component Regression* (PCR) untuk mengatasi multikolinearitas yang lebih baik?

Tujuan dari penelitian ini adalah Mengetahui metode yang lebih baik antara *Partial Least Square* (PLS) dan *Principal Component Regression* (PCR) untuk mengatasi multikolinearitas berdasarkan pada perbandingan nilai MSE dan ( $R^2$ ).

Untuk memperkirakan/meramalkan nilai variabel  $Y$ , akan lebih baik apabila diikuti memperhitungkan variabel-variabel lain yang ikut mempengaruhi  $Y$ . Dengan demikian, mempunyai hubungan diantara satu variabel tidak bebas (*dependent*)  $Y$  dengan beberapa variabel lain yang bebas (*independent*)  $X_1, X_2, \dots, X_p$ .

Untuk meramalkan  $Y$ , apabila semua nilai variabel bebas diketahui, maka kita dapat mempergunakan persamaan regresi linear berganda. Hubungan  $Y$  dan  $X_1, X_2, \dots, X_p$  yang sebenarnya adalah sebagai berikut.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon_1 \quad (1)$$

Istilah multikolinearitas mula-mula ditemukan oleh Ragnar Frisch. Pada mulanya multikolinearitas berarti adanya hubungan linear yang sempurna atau pasti, diantara beberapa atau semua variabel bebas dari model regresi ganda (Gujarati, 2004:157).

Ada beberapa cara untuk mengetahui keberadaan multikolinearitas dalam suatu model regresi (1) menganalisis matriks korelasi Jika antara dua atau lebih variabel *independent* memiliki korelasi yang cukup tinggi, biasanya di atas 0,9 maka hal tersebut mengindikasikan terjadinya multikolinearitas. (2) VIF (*Variance Inflation Factor*) adalah salah satu cara dalam mendeteksi adanya multikolinearitas, dan dalam penulisan ini menggunakan nilai Tolerance atau VIF. Multikolinearitas dalam sebuah regresi

dapat diketahui apabila nilai  $VIF \geq 10$ . (3) TOL (*Tolerance*) Jika nilai Tolerance kurang dari 0,1 atau nilai VIF melebihi 10 maka hal tersebut menunjukkan bahwa multikolinearitas adalah masalah yang pasti terjadi antar variabel bebas.

*Principal Component Regression* (PCR) merupakan salah satu metode yang telah dikembangkan untuk mengatasi masalah multikolinearitas. PCR merupakan analisis regresi dari variabel-variabel *dependent* terhadap komponen-komponen utama yang tidak saling berkorelasi, dimana setiap komponen utama merupakan kombinasi linear dari semua variabel *independent* (Draper & Smith 1992, hal 313).

*Principal Component Regression* (PCR) merupakan suatu teknik analisis yang mengkombinasikan antara analisis regresi dengan *Principal Component Analysis* (PCA). Analisis Regresi digunakan untuk mengetahui ada tidaknya hubungan antara variabel *dependent* dan *independent*, sedangkan PCA pada dasarnya bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensinya. Hal ini dilakukan dengan jalan menghilangkan korelasi di antara variabel melalui transformasi variabel asal ke variabel baru (merupakan kombinasi linear dari variabel-variabel asal) yang tidak saling berkorelasi. Dari  $p$  buah variabel asal dapat dibentuk  $p$  buah komponen utama, dipilih  $k$  buah komponen utama saja ( $k < p$ ) yang telah mampu menerangkan keragaman data cukup tinggi (antara 80% sampai dengan 90%) (Johnson & Wichern, 2010, hal. 356). Komponen utama yang dipilih tersebut ( $k$  buah) dapat mengganti  $p$  buah variabel asal tanpa mengurangi informasi.

Analisis regresi komponen utama (PCR) merupakan analisis regresi variabel *dependent* terhadap komponen-komponen utama yang tidak saling berkorelasi, regresi komponen utama dapat dinyatakan sebagai berikut :

$$Y = w_0 + w_1 K_1 + w_2 K_2 + \dots + w_m K_m + \varepsilon \quad (2)$$

$K_1, K_2, K_3, \dots, K_m$  menunjukkan komponen utama yang dilibatkan dalam analisis regresi komponen utama, dimana besaran  $m$  lebih kecil daripada banyaknya variabel *independent* yaitu sejumlah  $p$ , serta  $Y$  sebagai variabel *dependent*. Komponen utama merupakan kombinasi linear dari variabel baku  $Z$ , sehingga

$$K_1 = a_{11} Z_1 + a_{21} Z_2 + \dots + a_{p1} Z_p$$

$$K_2 = a_{12} Z_1 + a_{22} Z_2 + \dots + a_{p2} Z_p$$

⋮

$$K_m = a_{1m} Z_1 + a_{2m} Z_2 + \dots + a_{pm} Z_p \quad (4)$$

Apabila  $K_1, K_2, \dots, K_m$  dalam persamaan (4) didistribusikan kembali ke dalam persamaan regresi komponen utama, yaitu persamaan (3) maka diperoleh:

$$Y = w_0 + w_1(a_{11}Z_1 + a_{21}Z_2 + \dots + a_{p1}Z_p) + w_2(a_{12}Z_1 + a_{22}Z_2 + \dots + a_{p2}Z_p) + \dots + w_m(a_{1m}Z_1 + a_{2m}Z_2 + \dots + a_{pm}Z_p) + \varepsilon \quad (5)$$

Persamaan regresi linear dugaan komponen utama sebagai berikut:

$$Y = b_0 + b_1Z_1 + B_2Z_2 + \dots + b_pZ_p \quad (6)$$

Metode *Partial Least Square* (PLS) pertama kali diperkenalkan oleh Herman Ole Andres Wold pada tahun 1960 sebagai metode alternative untuk mengatasi keterbatasan metode *Ordinary Least Square* (OLS) ketika data mengalami masalah multikolinearitas. Untuk meregresikan variabel terikat  $y$  dengan variabel  $X_1, X_2, \dots, X_k$ , metode PLS mencari komponen-komponen baru yang berperan sebagai variabel bebas untuk mengestimasi parameter regresi.

Regresi PLS ini dapat diperoleh melalui regresi sederhana maupun berganda dengan mengambil kesimpulan dari uji signifikansi. Uji signifikansi ini bertujuan untuk memilih variabel *independent* pembangun komponen PLS dan menentukan banyaknya komponen PLS yang terbentuk (Vinzi *et al.*, 2004). Tujuan PLS adalah membentuk komponen yang dapat menangkap informasi dari variabel *independent* untuk memprediksi variabel *dependent*. Dalam pembentukan komponen PLS, digunakan variabel *dependent*  $y$  yang distandarisasi dan variabel-variabel *independent* yang terpusat (Vinzi, *et al.*, 2004). Model regresi *Partial Least Square* dengan  $m$  komponen dapat dituliskan sebagai berikut:

$$Y = \sum_{h=1}^m c_h t_h + \varepsilon \quad (7)$$

Dengan  $Y$  adalah variabel *dependent*,  $c_h$  adalah koefisien regresi  $Y$  terhadap  $t_h$ ,  $t_h = \sum_{j=1}^p w_{(h)j} X_j$  adalah komponen utama ke- $h$  yang tidak saling berkorelasi, ( $h=1,2,\dots,m$ ) dengan syarat komponen PLS  $t_h = \sum_{j=1}^p w_{(h)j} X_j$  orthogonal.

Penghitungan Komponen PLS pertama  $t_1$ . Komponen PLS pertama ( $t_1$ ) adalah linear dari variabel *independent*  $X_j$  dengan koefisien pembobot  $w_1$ .

$$t_1 = \frac{1}{\sqrt{\sum_{j=1}^p \text{cor}(x_j, y)^2}} \sum_{j=1}^p \text{cor}(x_j, y) x_j^* \quad (8)$$

Dimana  $x_j^*$  adalah  $x_j$  yang terstandarisasi,  $\text{cor}(x_j, y)$  adalah korelasi variabel  $x_j$  dengan  $y$ .

Perhitungan komponen PLS kedua  $t_2$ . Komponen PLS kedua didapatkan dengan melakukan regresi sederhana  $y$  terhadap  $t_1$  dan masing-masing  $x_j$  terlebih dahulu kemudian regresi antara  $x_j$  terhadap  $t_1$ . Variabel-variabel  $x_j$  yang digunakan hanya variabel yang berkontribusi secara signifikansi dalam menjelaskan  $y$  pada  $t_1$ . Model persamaan regresi keduanya adalah sebagai berikut:

$$t_2 = \frac{1}{\sqrt{\sum_{j=1}^p \text{cor}(y, x_{1j})^2}} \sum_{j=1}^p \text{cor}(y, x_{1j}) x_{1j} * \quad (9)$$

Dengan  $x_{1j} *$  adalah residu yang telah distandarisasi dan dihasilkan dari regresi  $x_j$  terhadap  $t_1$ .

Perhitungan Komponen PLS ke- $h$ ,  $t_h$  Seperti langkah pada pembentukan komponen PLS sebelumnya, variabel yang digunakan adalah variabel-variabel yang signifikan dalam menjelaskan  $y$  pada  $t_1, t_2, \dots, t_{h-1}$ . Model regresi  $y$  terhadap  $t_1, t_2, \dots, t_{h-1}$  dan masing-masing  $x_j$  adalah sebagai berikut:

$$y = c_1 t_1 + c_2 t_2 + \dots + c_{h-1} t_{h-1} + a_{hj} x_j + \text{residu}$$

Untuk mendapatkan komponen  $t_h$  yang orthogonal terhadap  $t_{h-1}$ , diregresikan  $x_j$  terhadap komponen PLS yang dituliskan sebagai berikut:

$$x_j = p_{1j} t_1 + p_{2j} t_2 + \dots + p_{h-1j} t_{h-1} + x_{(h-1)j}$$

Dengan  $x_{(h-1)j}$  adalah residu yang dihasilkan dari regresi setiap  $x_j$  terhadap  $t_1, t_2, \dots, t_{h-1}$ . Komponen ke- $h$  didefinisikan sebagai berikut:

$$t_h = \frac{1}{\sqrt{\sum_{j=1}^p \text{cor}(x_{(h-1)j}, y)^2}} \sum_{j=1}^p \text{cor}(x_{(h-1)j}, y) x_{(h-1)j}^* \quad (10)$$

Dengan  $a_{hj}$  adalah koefisien regresi dari  $x_{(h-1)j}$  dalam regresi  $y$  pada  $t_1, t_2, \dots, t_{h-1}$ .

Dimana  $x_{(h-1)j}^*$  adalah residu standar dari regresi dan setiap  $x_j$  terhadap  $t_1, t_2, \dots, t_{h-1}$ . Perhitungan komponen PLS berhenti ketika tidak ada lagi variabel *independent* yang signifikan membangun komponen PLS.

## METODE

Metode yang digunakan pada penelitian ini adalah fokus penelitian, studi pustaka, perumusan masalah, pengumpulan data, pemecahan masalah, dan menarik kesimpulan.

Fokus penelitian yang dilakukan dalam penelitian ini yaitu: (1) Penelitian menggunakan Model *Partial Least Square* (PLS), dan *Principal Component Regression* (PCR). (2) Metode yang digunakan adalah Metode *Partial Least Square* (PLS), dan *Principal Component Regression* (PCR). (3) Penelitian didukung dengan bantuan program SAS (*Statistical Analysis System*). (4) Variabel yang digunakan adalah Anggaran Pendapatan Daerah Provinsi Jawa Tengah 2013.

Studi pustaka dilakukan dengan cara mengumpulkan sumber-sumber informasi yang berkaitan dengan penelitian yang dikaji berupa buku, teks, jurnal, prosiding, dan literatur lainnya. Setelah sumber pustaka terkumpul, dilanjutkan dengan menelaah sumber pustaka tersebut untuk dijadikan landasan dalam analisis permasalahan yang dikaji.

Perumusan masalah diperlukan untuk membatasi permasalahan sehingga diperoleh bahan kajian yang jelas. Sehingga akan lebih mudah untuk menentukan langkah dalam memecahkan masalah tersebut. Berdasarkan latar belakang pemilihan judul dan materi pendukung yang telah dikaji, maka rumusan masalah pada penelitian ini yaitu (1) bagaimana model persamaan dengan metode *Partial Least Square* (PLS) untuk mengatasi multikolinearitas? (2) Bagaimana model persamaan dengan metode *Principal Component Regression* (PCR) untuk mengatasi multikolinearitas? (3) Metode manakah antara *Partial Least Square* (PLS) dan *Principal Component Regression* (PCR) untuk mengatasi multikolinearitas yang lebih baik?

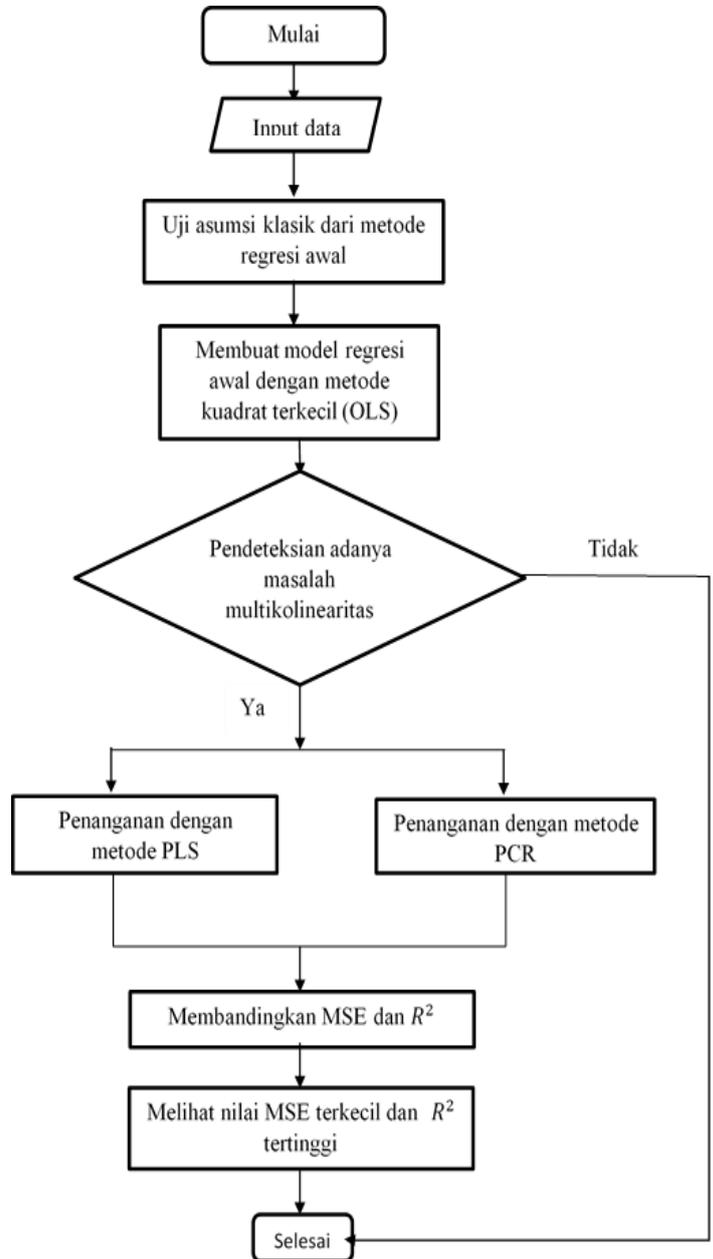
Pengumpulan data dengan cara mengambil data sekunder yang diperoleh dari Kantor Badan Pusat Statistik (BPS) Provinsi Jawa Tengah, data yang dipakai Anggaran Pendapatan Daerah Provinsi Jawa Tengah 2013.

Pada tahap ini dilakukan kajian pustaka, yaitu mengkaji permasalahan secara teoritis berdasarkan sumber-sumber pustaka yang ada. Adapun langkah-langkah yang dilakukan dalam tahap pemecahan masalah ini adalah :

1. Mencari data yang akan digunakan.
2. Menentukan variabel terikat dan variabel bebas dari data yang diperoleh.
3. Melakukan analisis regresi untuk menentukan model regresi dengan metode kuadrat terkecil.
4. Melakukan analisis regresi untuk melihat ada tidaknya multikolinearitas dengan melihat nilai VIF dan TOL.
5. Melakukan penanganan terhadap masalah multikolinearitas apabila asumsi multikolinearitas terpenuhi.

6. Membandingkan nilai koefisien determinasi  $R^2$  dan nilai *Mean Square Error* (MSE) untuk memperoleh metode terbaik.
7. Menggunakan metode terbaik yang didapat untuk mengestimasi data.

Langkah-langkah tersebut dapat dilihat pada diagram alur Gambar 1 sebagai berikut.



Gambar 1 Diagram Alur Penelitian

Penarikan kesimpulan merupakan langkah terakhir dalam kegiatan penelitian ini yang dilakukan dengan cara menarik kesimpulan dari keseluruhan permasalahan yang telah dirumuskan dengan berdasarkan pada landasan teori dan hasil pemecahan masalah.

**HASIL DAN PEMBAHASAN**

Dalam penelitian ini, langkah pertama yang dilakukan adalah mendeskripsikan data Anggaran Pendapatan Daerah Provinsi Jawa Tengah Tahun 2013. Selanjutnya akan dilakukan pengujian asumsi regresi, uji asumsi klasik, mengatasi masalah multikolinearitas dan pemilihan metode terbaik.

1. Uji Asumsi Regresi
  - a. Uji Normalitas

Model regresi yang baik jika distribusi data normal atau mendekati normal. Hal ini dapat dideteksi dengan melihat nilai *Kolmogorov Smirnov* (K-S). berikut nilai *Kolmogorov Smirnov* (K-S). Plot Grafik Normal  $p - p$  .

Tabel 1. Normalitas Data

Goodness-of-Fit Tests for Normal Distribution				
test	statistik	DF	P Value	
Kolmogorov-Smirnov	D	0,136586	0,095	
Cramer-von Mises	W-Sq	0,157733	0,019	
Anderson-Darling	A-Sq	0,88332	0,022	
Chi-Square	Chi-sq	14,64407	4	0,005

Output SAS nilai Kolmogorov-Smirnov menunjukkan 0,1365 dan P value 0,095. Hal ini menunjukkan bahwa data berdistribusi normal.

- b. Model Regresi Berganda

Dari data Anggaran Pendapatan Daerah Provinsi Jawa Tengah dilakukan regresi linear dengan menggunakan program SAS. Analisis regresi bertujuan untuk mengetahui hubungan variabel-variabel *dependent* terhadap variabel *independentnya*.

Tabel 2. Parameter Estimates

Parameter Estimates			
Variable	Parameter estimate	Standard error	t Value
	115519965	39740014	2,91
X1	-0,86615	1,03329	-0,84
X2	3,60648	3,6006	1
X3	-2,6309	2,38674	-1,1
X4	1,6646	0,40515	4,11
X5	0,95078	0,03853	24,68
x6	1,0347	0,08472	12,21

Berdasarkan tabel 2 hasil persamaan regresi linear dugaan yang diperoleh adalah sebagai berikut:

$$\hat{y} = 115519965 - 0,866615X_1 + 3,60648X_2 - 2,63090X_3 + 1,66460X_4 + 0,95078X_5 + 1,03470X_6$$

2. Uji Asumsi Klasik

Uji asumsi klasik adalah persyaratan statistik yang harus dipenuhi pada analisis regresi linear berganda. Uji asumsi klasik yang digunakan yaitu uji multikolinearitas, uji heteroskedastisitas dan uji autokorelasi dengan menggunakan program SAS sehingga diperoleh hasil sebagai berikut:

- a. Multikolinearitas

Uji multikolinearitas bertujuan untuk menguji apakah dalam model regresi terdapat korelasi antar variabel *independent*. Teknik yang dapat digunakan untuk mendeteksi adanya multikolinearitas diantaranya adalah pemeriksaan matriks korelasi, nilai VIF dan TOL tapi dalam penelitian ini peneliti hanya melihat nilai VIF dan TOL untuk mendeteksi adanya multikolinearitas pada data. Berikut nilai *Variance Inflation Factor* (VIF) dan TOL.

Tabel 3. Nilai VIF dan TOL

Parameter Estimates		
Variable	TOL	VIF
Intercept	.	0
X1	0,24771	4,03702
X2	0,59049	1,69350
X3	0,06067	16,48238
X4	0,06079	16,45103
X5	0,65004	1,53837
X6	0,85208	1,17360

Variabel  $X_3$  dan  $X_4$  mengandung multikolinearitas sebab dari nilai VIF > 10 yaitu 16,48238 dan 16,45103 dan nilai TOL < 0,1 yaitu 0,06067 dan 0,06079. sehingga akan dilakukan tindakan lanjut untuk mengatasi masalah multikolinearitas dengan menggunakan metode PCR dan PLS.

- b. Autokorelasi

Hipotesis:

$H_0$  : Tidak terjadi autokorelasi

$H_0$  : Terjadi autokorelasi

Tingkat signifikansi  $\alpha = 0,05$

Kriteria:

Jika  $d > d_u$ , maka  $H_0$  diterima (tidak ada autokorelasi).

Jika  $d < d_L$ , maka  $H_0$  ditolak (ada autokorelasi positif).

Jika  $4 - d > d_u$ , maka  $H_0$  diterima (tidak ada autokorelasi negatif).

Jika  $4 - d < d_L$ , maka  $H_0$  ditolak (ada autokorelasi negatif).

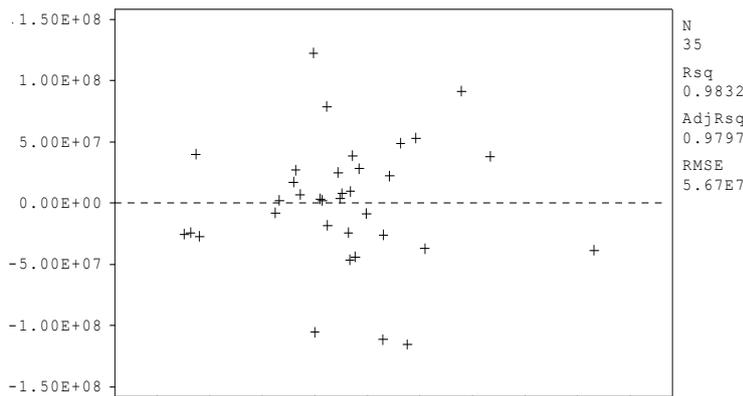
Tabel 4. Autokorelasi

Durbin-Watson D	2,181
Number of Observations	35

Berdasarkan output SAS di atas diperoleh nilai Durbin-Watson D sebesar 2,181. Langkah selanjutnya menentukan nilai  $d_L$  dan  $d_u$  dengan taraf nyata 0,05, observasi sebanyak  $n = 35$  data, dan K adalah jumlah variabel *independent*nya sebanyak 6, dengan melihat nilai tabel Durbin-Watson (dengan Taraf signifikansi 0,05) maka diperoleh nilai  $d_L = 1,097$  sedangkan nilai  $d_u = 1,884$ . Karena  $d = 2,181 > d_u = 1,884$  maka  $H_0$  diterima (tidak ada autokorelasi).

c. Uji Heteroskedastisitas

Untuk mendeteksi ada tidaknya heteroskedastisitas dengan melihat pola titik-titik pada plots regresi. Jika titik-titik menyebar dengan pola tidak jelas di atas dan di bawah angka 0 pada sumbu Y maka tidak terjadi masalah heteroskedastisitas. Berikut output plot dari program SAS terlihat pada gambar 2.



Gambar 2. Plot Heteroskedastisitas.

Gambar 2 dapat diketahui bahwa titik-titik menyebar dengan pola yang tidak jelas di atas

dan di bawah angka 0 pada sumbu Y maka tidak terjadi masalah heteroskedastisitas.

3. *Principal Component Regression* (PCR)

a. Menentukan Komponen Utama

*Principal Component Regression* (PCR) merupakan teknik analisis regresi yang dikombinasikan dengan teknik analisis komponen utama, dimana analisis komponen utama dijadikan sebagai tahap analisis. Oleh karena itu, sebelum melakukan PCR terlebih dahulu dilakukan analisis komponen utama untuk mendapatkan komponen-komponen utama dan skor komponen utama yang berguna sebagai variabel-variabel *independent* dalam PCR. Langkah pertama adalah dengan melihat nilai Commuality yang menunjukkan berapa varians yang dapat dijelaskan oleh komponen yang terbentuk. Hasil varians tersebut dapat ditunjukkan dalam tabel di bawah ini:

Tabel 5. Final Commuality Estimates

Final Commuality Estimates					
X1	X2	X3	X4	X5	X6
0,880	0,812	0,924	0,936	0,904	0,719

Hasil yang diperoleh dari tabel 5 memperlihatkan bahwa nilai variabel  $X_1$  sebesar 0,88 yang berarti sekitar 88 % variansi variabel  $X_1$  dapat dijelaskan oleh komponen yang terbentuk, variabel  $X_2$  sebesar 0,811 yang berarti 81,1 % variansi variabel  $X_2$  dijelaskan oleh komponen yang terbentuk, variabel  $X_3$  sebesar 0,924 yang berarti 92,4 % variansi variabel  $X_3$

dijelaskan oleh komponen yang terbentuk, variabel  $X_4$  sebesar 0,936 yang berarti 93,6 % variansi variabel  $X_4$  dijelaskan oleh komponen yang terbentuk, variabel  $X_5$  sebesar 0,904 yang berarti 90,4 % variansi variabel  $X_5$  dijelaskan

oleh komponen yang terbentuk, dan variabel  $X_6$  sebesar 0,719 yang berarti 71,9 % variansi variabel  $X_6$  dijelaskan oleh komponen yang terbentuk.

Langkah selanjutnya adalah *Eigenvalues of the Correlation Matrix* dalam analisis faktor terdapat beberapa komponen yang merupakan variabel. Setiap faktor mewakili variabel yang dianalisis. Kemampuan setiap faktor mewakili variabel yang dianalisis ditunjukkan oleh besarnya varians yang dijelaskan, yang disebut dengan *eigenvalue*. *Eigenvalue* menunjukkan kepentingan relatif masing-masing faktor dalam menghitung varians semua variabel yang dianalisis. Susunan *eigenvalue* selalu diurutkan dari yang terbesar sampai yang terkecil, dengan kriteria bahwa angka *eigenvalue* di bawah 1 tidak digunakan dalam menghitung jumlah faktor yang terbentuk.

Tabel 6. Eigenvalues of the Correlation Matrix

Eigenvalues of the Correlation Matrix			
	Eigenvalue	Difference	Cumulative
1	3,00854913	1,63430587	0,5014
2	1,37424326	0,57267512	0,7305
3	0,80156814	0,19483842	0,8641
4	0,60672971	0,42928168	0,9652
5	0,17744804	0,1459863	0,9948
6	0,03146173		1

Pada tabel 6 terlihat bahwa angka *eigenvalues* di bawah 1 tidak dapat digunakan dalam menghitung jumlah faktor yang terbentuk, sehingga proses factoring seharusnya berhenti pada dua faktor saja.

*Eigenvalue* di variabel  $X_1$  sebesar 3,008, artinya faktor ini dapat menjelaskan 50,14 % dari total cumulative dan variabel  $X_2$  sebesar 1,374 yang artinya faktor ini menjelaskan 73,05% dari total *cumulative*.

Langkah selanjutnya adalah dengan mencari nilai skor komponen yang terbentuk untuk membentuk persamaan regresi komponen utama. skor komponen ini nantinya akan menentukan banyaknya variabel yang akan terbentuk untuk menggantikan variabel-variabel  $X_1, X_2, X_3, X_4, X_5$  dan  $X_6$  dengan variabel baru.

Tabel 7. Skor Komponen Utama

	Scoring Coefficients factor	
	Factor1	Factor2
X1	0,28599	0,14405
X2	0,03591	0,56114
X3	0,3268	-0,1536
X4	0,33003	-0,0912
X5	0,18526	0,07292
X6	-0,0775	0,58999

Berdasar tabel 7, maka persamaan regresi komponen utama yang dapat dibentuk adalah sebagai berikut.

$$K_1 = 0,28599 X_1 + 0,03591 X_2 + 0,32680 X_3 + 0,33003 X_4 + 0,18526 X_5 - 0,07747 X_6$$

$$K_2 = 0,14405 X_1 + 0,56114 X_2 - 0,15359 X_3 - 0,09120 X_4 + 0,07292 X_5 + 0,58999 X_6$$

Setelah skor komponen terbentuk langkah selanjutnya adalah menentukan variabel-variabel baru untuk menggantikan variabel-variabel  $X_1, X_2, X_3, X_4, X_5$  dan  $X_6$ . Pembentukan komponen yang terbentuk dengan cara menstandarisasikan semua variabel dan mengalikan dengan skor komponen utama. Maka variabel baru yang terbentuk seperti tabel 8.

Variabel-variabel tersebut menunjukan bahwa dua komponen adalah jumlah yang paling optimal untuk mereduksi ke-enam variabel-variabel *independent*.

b. Regresi Komponen Utama

Hubungan antara variabel-variabel *dependent* dan variabel *independent* dapat diketahui dengan melakukan analisis *Principal Component Regression* (PCR). Dalam hal ini, variabel *dependent* (Y) anggaran pendapatan daerah Provinsi Jawa Tengah diregresikan dengan variabel baru, yaitu variabel komponen utama yang terbentuk. Dengan demikian, model regresi komponen utama yang dibangun untuk kasus ini adalah :

$$Y = W_0 + W_1 K_1 + W_2 K_2 + \varepsilon$$

Maka model persamaan persamaan yang terbentuk pada tabel 9.

Tabel 8. Variabel Baru

Factor 1	Factor 2
0,351735	1,432907
0,337589	1,587791
-0,24977	0,557758
-0,4089	-0,01168
0,138951	-0,89797
-0,48721	-0,30175
-0,31339	-0,59496
0,114003	0,693008
0,011316	0,768867
-0,23544	0,784044
-0,03959	-0,30474
-0,02198	-0,71316
0,766953	2,901944
-0,26602	0,633214
-0,33751	1,115437
-0,59022	0,071928
-0,14045	-1,11048
-0,23063	1,051383
-0,35497	-0,27354
-0,39793	0,156588
-0,44103	0,41592
-0,06746	0,357698
-0,49773	0,331697
0,160377	-0,69028
-0,19881	-0,27498
-0,52837	-1,10275
-0,25125	0,898536
0,01058	-0,37572
0,126957	-1,23403
-0,77816	-1,08327
1,337884	0,483484
-0,70272	-1,31915
5,22719	-1,19602
-0,53571	-1,24638
-0,50829	-1,51134

Tabel 9 Regresi Y Terhadap Komponen Utama  
Parameter Estimates

Variable	Parameter Estimate	Standard error	TOL	VIF
Intercept	1382126508	45674516	.	0
K1	258988114	46341335	1	1
K2	149127143	46341346	1	1

$$\hat{Y} = 1382126508 + 258988114K_1 + 149127143K_2$$

$$\hat{Y} = 1382126508 + 258988114 (0,28599 X_1 + 0,03591 X_2 + 0,32680 X_3 + 0,33003 X_4 + 0,18526 X_5 - 0,07747X_6) + 149127143 (0,14405 X_1 + 0,56114 X_2 - 0,15359 X_3 - 0,09120 X_4 + 0,07292 X_5 + 0,58999X_6)$$

$$\bar{Y} = 1382126508 + 28888566,77X_1 + 84611231,43X_2 - 14440705,5X_3 - 5053009,86X_4 - 6076336,99X_5 - 89989904,2X_6$$

Setelah diperoleh hasil regresi linear dugaannya, maka dilakukan pengecekan apakah data masih mengandung multikolinearitas. Hal ini dapat dilihat pada tabel 9 bahwa nilai VIF=TOL=1 yang berarti tidak terjadi multikolinearitas.

4. *Partial Least Square* (PLS)

a. Pembentukan Komponen PLS pertama

Sebelum pembentukan komponen PLS pertama, terlebih dahulu dilakukan regresi y terhadap masing-masing x untuk mengetahui variabel-variabel manakah yang signifikan membangun komponen PLS pertama. Berikut adalah hasil regresi y (terstandarisasi) terhadap masing-masing x.

Tabel 10 Uji Signifikansi

Parameter Estimates			
Variable	Parameter Estimate	Standard Error	t Value
X1	3,98E-08	9,57E-09	4,16
X2	1,36E-07	6,38E-08	2,12
X3	2,89E-08	9,24E-09	3,13
X4	6,24E-09	1,42E-09	4,39
X5	2,83E-09	2,59E-10	10,93
X6	3,15E-09	1,29E-09	2,45

Uji signifikansi koefisien regresi pada Tabel 10 menunjukkan bahwa dengan taraf nyata 5% semua variabel signifikan membangun komponen PLS pertama.

Pengambilan keputusan:

T hitung  $\leq$  t kritis jadi  $H_0$  diterima

T hitung  $>$  t kritis jadi  $H_1$  ditolak.

Titik kritis dapat dicari pada tabel statistik pada signifikansi dengan  $df = n - k - 1$  maka  $df = 35 - 6 - 1 = 29$  (k adalah jumlah variabel *independent*). Di dapat t kritis adalah 2,04523.

$$t_1 = \frac{1}{\sqrt{\sum_{j=1}^p \text{cor}(x_j, y)^2}} \sum_{j=1}^p \text{cor}(x_j, y)x_j^*$$

Dimana  $x_j^*$  adalah variabel *independent* yang telah distandarisasi, sehingga komponen PLS pertama yang terbentuk.

$$t_1 = 0,385X_1^* + 0,184X_2^* + 0,340X_3^* + 0,434X_4^* + 0,665X_5^* + 0,271X_6^*$$

Maka diperoleh nilai komponen PLS pertama

Tabel 11. Komponen PLS 1

PLS1
1,903781
1,765623
-0,24876
-0,52369
0,266509
-0,78432
-0,69573
0,524041
0,712924
0,544488
-0,36787
-0,04235
-0,10292
0,069222
0,23434
-0,75894
-0,73036
0,572452
-0,49633
-0,368
-0,32267
0,060653
-0,88761
0,114955
-0,59971
-1,3015
0,232918
0,375152
0,315051
-2,29032
1,947366
-2,29342
7,132047
-1,94483
-2,01218

b. Pembentukan Komponen PLS kedua

Sebelum pembentukan komponen PLS kedua, terlebih dahulu diperiksa apakah komponen kedua ini masih diperlukan. Hal tersebut dilakukan dengan cara meregresikan antara  $y$  yang telah distandarisasi terhadap  $t_1$  dan masing-masing variabel  $x$ . Variabel yang digunakan adalah variabel-variabel yang signifikan membangun PLS kedua.

Tabel 12 Uji Signifikansi

Parameter Estimates			
Variable	Parameter Estimate	Standard Error	t Value
X1	-2,04E-08	6,21E-09	-3,28
X2	5,47E-09	2,49E-08	0,22
X3	-3,91E-08	5,13E-09	-7,62
X4	-6,14E-09	1,34E-09	-4,58
X5	1,65E-09	2,94E-10	-5,61
X6	1,77E-09	6,21E-10	2,01

Berdasarkan tabel 12, ternyata tidak ada variabel yang signifikan yaitu  $X_1, X_2, X_3, X_4, X_5$  dan  $X_6 <$  dari t kritis 2,04523, sehingga perhitungan berhenti pada komponen PLS pertama. Setelah mendapatkan komponen baru pada tabel 11, kemudian variabel  $y$  diregresikan terhadap komponen PLS yang terbentuk. Berikut hasil regresi yang diperoleh:

Tabel 13 Regresi Y Terhadap Variabel PLS

Parameter Estimates				
Variable	Parameter Estimate	Standard error	TOL	VIF
Intercept	1382126382	32340470	.	0
PLS1	227404292	30760787	1	1

$$\hat{Y} = 1382126382 + 218046820 \text{ PLS1} \hat{Y}$$

$$\hat{Y} = 1382126382 + 218046820 (0,385X_1^* + 0,184X_2^* + 0,340X_3^* + 0,434X_4^* + 0,665X_5^* + 0,271X_6^*)$$

$$\hat{Y} = 1382126382 + 83948025,7X_1 + 40120614,88X_2 + 74135918,8X_3 + 94632319,88X_4 + 145001135,3X_5 + 59090688,22X_6$$

Setelah diperoleh hasil regresi linear dugaannya, maka dilakukan pengecekan apakah data masih mengandung multikolinearitas. Hal ini dapat dilihat pada tabel 13 bahwa nilai  $VIF=TOL=1$  yang berarti tidak terjadi multikolinearitas.

#### 5. Pemilihan Metode Terbaik

Pemilihan model terbaik untuk mengatasi masalah multikolinearitas PCR dan PLS adalah membandingkan hasil persamaan regresi yang diperoleh antara kedua metode tersebut. Koefisien determinasi  $R^2$  dapat digunakan untuk menentukan model terbaik. Semakin nilai  $R^2$  mendekati satu maka semakin tinggi pengaruh variabel *independent* terhadap variabel *dependent*, yang berarti semakin baik kecocokan model dengan data. Selain melihat nilai  $R^2$  pemilihan model terbaik juga dapat dilakukan dengan melihat nilai *Mean Square Error* (MSE). Metode terbaik adalah metode dengan nilai MSE terkecil. Perbandingan nilai  $R^2$  dan MSE dari metode PLS dan PCR adalah sebagai berikut:

Tabel 14 Nilai  $R^2$  dan MSE

Metode	$R^2$	MSE
PLS	0,7752	3,660671E16
PCR	0,5652	7,301565E16

Berdasarkan tabel 14 dapat dilihat bahwa nilai  $R^2$  yang dihasilkan PLS dengan nilai  $R^2 = 0,7752$  lebih besar dibandingkan PCR dengan nilai  $R^2 = 0,5652$  dan nilai MSE yang dihasilkan PLS sebesar 3,660671E16 lebih kecil dibandingkan dengan nilai MSE PCR sebesar 7,301565E16 maka dapat disimpulkan bahwa metode PLS lebih baik dibandingkan dengan metode PCR dalam mengatasi masalah multikolinearitas.

#### SIMPULAN

Simpulan yang diperoleh dari penelitian ini adalah (1) Model persamaan dengan metode *Partial Least Square* (PLS) pada kasus pendapatan anggaran Daerah Provinsi Jawa Tengah sebagai berikut.

$$\hat{Y} = 1382126382 + 83948025,7X_1 + 40120614,88X_2 + 74135918,8X_3 + 94632319,88X_4 +$$

145001135,3X<sub>5</sub> + 59090688,22X<sub>6</sub> (2) Model persamaan dengan metode *Principal Component Regression* (PCR) pada kasus pendapatan anggaran Daerah Provinsi Jawa Tengah sebagai berikut.

$$\hat{Y} = 1382126508 + 28888566,77X_1 + 84611231,43X_2$$

$$-14440705,5X_3 - 5053009,86X_4 - 6076336,99X_5 - 89989904,2X_6 \text{ (3)}$$

Perbandingan metode manakah antara *Partial Least Square* (PLS) dan *Principal Component Regression* (PCR) untuk mengatasi multikolinearitas yang lebih baik dilihat dari nilai  $R^2$  dan nilai MSE. dapat dilihat bahwa nilai  $R^2$  yang dihasilkan PLS = 0,7752 lebih besar dibandingkan dengan PCR = 0,5652 dan nilai MSE yang dihasilkan PLS = 3,660671E16 lebih kecil dibandingkan dengan PCR = 7,301565E16 maka dapat disimpulkan bahwa metode PLS lebih baik dibandingkan dengan metode PCR dalam mengatasi masalah multikolinearitas.

#### DAFTAR PUSTAKA

- Abdi, H. 2010. *Partial Least Square (PLS) Regression and Projrction On Latent Structure Regression. Encyclopedia of Social Sciences Research Methods.*
- Draper, H., & Smith, H. 1992. *Applied Regression Analysis, 2nd. (Analisis Regresi Terapan Edisi Ke-2). Penerjemah : Bambang Sumantri. Jakarta:PT. Gramedia Pustaka Utama.*
- Farahani, A, Rahiminezhed H., Same A.L, & Immannezhed, K. 2010. A Comparison of Partial Least Square (PLS) and Ordinary Least Square (OLS) regressions in predicting of couples mental health based on their communicational patterns. *Procedia.*
- Gujarati, D. 2004. *Ekonometrika Dasar.* Sumarno Zain Penerjemah. Jakarta: Erlangga. Terjemahan dari: *Basic Econometrics.*
- Johnson, R.A., Wichern, D.W. 2010. *Applied Multivariate Statistical Analysis.* Sixth edition, Prentice Hall. New Jersey.
- Nurhasanah, Subianto, M. & Fitriani, R. 2012. Perbandingan Metode Partial Least Square (PLS) dengan Regresi Komponen Utama untuk Mengatasi Multikolinearitas. *Jurusan Matematika FMIPA UNSYIAH, Statistika, 12 (1): 33 – 42.*
- Saikia 2014. Estimation of Principal Components Regression Coefficients. *International of Science and Research (IJSR).*
- Soemartini. 2008. *Penyelesaian Multikolinearitas Melalui Ridge Regression.* Tesis. Universitas Padjajaran.
- Sukmono, A. 2014. *Penggunaan Partial Least Square Regression (PLSR) untuk Mengatasi Multikolinearitas Dalam Estimasi Klorofil Daun Tanaman Padi dengan Citra Hiperspektral.* Program Studi Teknik Geodesi, Fakultas Teknik, Universitas Diponegoro.

- Tazliqoh, A., H. Rahmawati, H., & Safitri, D. 2015. Perbandingan Regresi Komponen Utama Regresi Ridge Pada Analisis Faktor-Faktor Pendapatan Asli Daerah (PAD) Provinsi Jawa Tengah. *Jurnal Gaussian*.
- Widarjono, A. 2007. *Ekonometrika Teori dan Aplikasi untuk Ekonomi dan Bisnis*. Yogyakarta: Ekonisia FE UII.
- Vinzi, V., Bastien, P., & Tenenhaus, M. 2004. Partial Least Square Generalized Linear Regression. *Computational statistics & Data Analysis (2005)*.