

# **Model *Partial Least Square Regression* (PLSR) Pengaruh Bidang Pendidikan dan Ekonomi Terhadap Tingkat Kemiskinan di Indoneia**

**I Gusti Ayu Made Srinadi**

Jurusan Matematika, FMIPA, Universitas Udayana, Bukit Jimbaran, Badung (80361)

Email: [srinadi@unud.ac.id](mailto:srinadi@unud.ac.id)

**Abstract:** *Partial Least Square Regression (PLSR) is one of the methods applied in the estimation of multiple linear regression models when the ordinary least square method (OLS) can not be used. OLS generates an invalid model estimate when multicollinearity occurs or when the number of independent variables is greater than the number of data observations. In conditions that OLS can be applied in obtaining model estimation, want to know the performance of PLSR method. This study aims to determine the model of PLSR the influence of literacy rate, the average of school duration, school enrollment rate, Income per capita, and open unemployment rate to the level of poverty seen from the percentage of poor people in Indonesia by 2015. Estimated model with OLS, Only variable of literacy rate are included in the model with the coefficient of determination  $R^2 = 32.52\%$ . PLSR model estimation of cross-validation, leave-one-out method with one selected component has  $R^2$  of 33,23%. Both models shows a negative relationship between poverty and literacy rate. The higher literacy rate will reduce the poverty level, indicating that the success of the Indonesian government in the development of education will support the government's success in reducing poverty level.*

**Keywords:** *multicollinearity, cross-validation, leave-one-out, component, coefficient of determination*

## **1. Pendahuluan**

Model regresi dalam analisis regresi linear berganda dapat digunakan sebagai model prediksi yang bersifat interpolasi. Prediksi bersifat interpolasi dimaksudkan adalah menduga nilai suatu peubah respon pada nilai-nilai peubah bebas yang berada dalam interval nilai dari peubah-peubah bebas yang digunakan dalam pembentukan model. Model regresi linear berganda memuat satu peubah respon (Y) dan peubah bebasnya (X) lebih dari satu. Model estimasi dikatakan cukup baik menggambarkan keragaman data apabila koefisien determinasi ( $R^2$ ) model tersebut untuk bidang sains  $\geq 80\%$ , dan untuk bidang sosial nilai  $R^2 \geq 65\%$  [1]. Model regresi linear berganda yang diperoleh dengan menggunakan metode kuadrat terkecil (*ordinary least square-OLS*)

sering menjadi tidak valid apabila terjadi korelasi yang signifikan antar peubah bebas. Korelasi antar peubah bebas disebut sebagai kasus multikolinearitas dalam regresi linear berganda. PLSR merupakan salah satu metode yang dapat diterapkan dalam estimasi model regresi bila terjadi kasus multikolinearitas, tanpa mereduksi peubah yang masuk dalam model. PLSR diperkenalkan oleh Wold (1982) yang digunakan untuk membangun model yang bersifat prediksi [2]. PLSR merupakan analisis regresi bersifat tidak langsung dimana varians dari peubah respon dikaitkan dengan besarnya peubah bebas melalui satu atau beberapa faktor yang didefinisikan sebagai kombinasi linear dari peubah bebas yang disebut dengan komponen. PLSR merupakan salah satu metode yang didesain untuk menyelesaikan regresi linear berganda ketika terjadi permasalahan spesifik pada data, seperti ukuran sampel penelitian yang kecil (*microarray data*), adanya data yang hilang (*missing values*), terjadi multikolinearitas, dan dapat diterapkan pada tipe data yang berbeda-beda [3].

Penelitian ini bertujuan untuk menentukan model PLSR pengaruh angka melek huruf (AMH), rata-rata lama sekolah, angka partisipasi sekolah, pendapatan perkapita, dan tingkat pengangguran terbuka terhadap tingkat kemiskinan yang dilihat dari persentase penduduk miskin di Indonesia tahun 2015.

Keterkaitan kemiskinan dan pendidikan sangat besar karena pendidikan memberikan kemampuan untuk berkembang lewat penguasaan ilmu dan keterampilan. Pendidikan juga menanamkan kesadaran akan pentingnya martabat manusia. Mendidik dan memberikan pengetahuan berarti menggapai masa depan. Hal tersebut harusnya menjadi semangat untuk terus melakukan upaya mencerdaskan bangsa [4]. Penelitian yang dilakukan oleh Hermanto Siregar dan Dwi Wahyuniarti menemukan bahwa pendidikan berpengaruh negatif terhadap tingkat kemiskinan. Hal ini menunjukkan bahwa pendidikan sangat penting dalam menurunkan tingkat kemiskinan.

Faktor lain yang juga berpengaruh terhadap tingkat kemiskinan adalah pengangguran. Salah satu unsur yang menentukan kemakmuran suatu masyarakat adalah tingkat pendapatan. Pendapatan masyarakat mencapai maksimum apabila kondisi tingkat penggunaan tenaga kerja penuh (*full employment*) dapat terwujud. Pengangguran akan menimbulkan efek mengurangi pendapatan masyarakat, dan menurunkan tingkat kemakmuran yang telah tercapai. Semakin turun tingkat kemakmuran akan menimbulkan masalah lain yaitu kemiskinan [5].

Model regresi untuk mengetahui pengaruh angka melek huruf (AMH), rata-rata lama sekolah, angka partisipasi sekolah, pendapatan perkapita, dan tingkat pengangguran terbuka terhadap tingkat kemiskinan pada data amatan yang cukup dapat diestimasi dengan metode kuadrat terkecil. Secara matriks, model regresi linear berganda dinyatakan dalam bentuk:

$$Y = X\beta + \epsilon$$

dengan:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}_{n \times 1} ; X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}_{n \times p} ; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix}_{p \times 1} ; \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_1 \\ \dots \\ \epsilon_n \end{bmatrix}_{n \times 1} \text{ dengan } p=k+1.$$

Estimasi model regresi identik dengan estimasi parameter atau koefisien regresi ( $\beta_i$ ), untuk  $i = 0, 1, \dots, k$ , misalkan nilai dugaan parameter regresi dinyatakan dengan ( $b_i$ ) maka estimasi model regresi dinyatakan sebagai:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k = Xb$$

dengan  $\hat{\beta} = b = (X'X)^{-1}X'Y$ . Penduga parameter regresi dengan MKT merupakan penduga yang bersifat tidak bias apabila matriks  $X'X$  merupakan matriks tidak singular, yaitu matriks yang memiliki invers yang bersifat unik. Apabila  $X'X$  bersifat singular atau dalam kondisi buruk (*ill conditions*) yaitu nilai determinannya mendekati nol, maka penduga parameternya akan bersifat bias. Kondisi buruk disebabkan salah satunya karena terjadi multikolinearitas antar peubah bebas. Metode alternatif menentukan estimasi model dalam kondisi buruk salah satunya dengan PLSR.

Menurut Esposito [3], metode PLSR memiliki keunggulan dan kelemahan. Keunggulan dari PLSR adalah kemampuan dari model peubah berganda berganda untuk peubah respon tunggal sama baiknya dengan peubah respon berganda, kemampuan dalam menangani kasus multikolinearitas antar peubah bebas, ketahanan dalam menghadapi data yang hilang, dan membuat peubah laten (komponen utama) secara langsung berdasarkan *crossproduct* yang melibatkan peubah respon untuk membuat prediksi menjadi lebih kuat. Adapun kelemahan dari PLSR adalah kesulitan yang besar dalam menafsirkan *loading* pada peubah bebas laten (yang didasarkan pada hubungan *crossproduct* dengan peubah respon, tidak seperti pada analisis faktor berdasarkan matriks kovarians antar peubah bebas).

Komponen pertama pada model PLSR dipilih untuk memaksimalkan nilai kovarian dengan peubah respon dan kemudian digunakan sebagai *regressor* (peubah bebas) dalam model regresi pada MKT (*Ordinary Least Square – OLS*). Komponen selanjutnya dipilih untuk memaksimumkan nilai kovarian dengan sisaan pada penduga OLS. Hasil akhir dari model PLSR disajikan sebagai  $\hat{b}_{PLS}$  dari koefisien regresi, dengan model:

$$Y = \bar{Y} + b(X - \bar{X}) + \epsilon$$

dimana Y adalah vektor yang mengandung nilai-nilai respon, X adalah matriks dari peubah penjelas dan  $\epsilon$  adalah vektor dari kesalahan yang tidak berkorelasi dengan varians yang sama [6].

Banyaknya peubah laten yang digunakan pada akhir model PLSR dapat ditentukan dengan validasi silang, dengan mempelajari kemampuan model yang berbeda untuk memprediksi peubah respon. Dalam penelitian umumnya pengamatan dibagi menjadi dua set data, satu digunakan untuk kecocokan model dan lainnya untuk validasi model, selanjutnya dibandingkan nilai prediksi dan nilai observasi sebenarnya [7].

Pada kasus multikolinearitas, PLSR menghindari matriks  $\mathbf{X}\mathbf{X}$  yang bersifat singular dengan dekomposisi  $\mathbf{X}$  ke dalam orthogonal *scores*  $\mathbf{T}$  dan *loadings*  $\mathbf{P}$ .

$$\mathbf{X} = \mathbf{T}\mathbf{P}$$

dan meregresikan  $\mathbf{Y}$  tidak pada  $\mathbf{X}$  itu sendiri, tetapi dengan kolom pertama dari *scores*  $\mathbf{T}$ . PLSR bertujuan untuk menggabungkan informasi dari  $\mathbf{X}$  dan  $\mathbf{Y}$  yang didefinisikan oleh *scores* dan *loadings*. Dalam faktanya, untuk satu versi spesifik dari algoritma *Partial Least Square* dapat menunjukkan *scores* dan *loadings* yang dipilih sedemikian rupa untuk menggambarkan sebanyak mungkin kovarians diantara  $\mathbf{X}$  dan  $\mathbf{Y}$  [8].

Bila dimiliki  $k$  buah peubah bebas dan sebuah peubah respon, model PLSR dengan  $m$  komponen dapat juga dirumuskan sebagai [9]:

$$Y = \sum_{h=1}^m c_h \left( \sum_{j=1}^p w_{hj}^* X_j \right) + \epsilon$$

Komponen pertama dinyatakan dengan  $t_1$ , didefinisikan sebagai:

$$t_1 = Xw_1^* = \frac{1}{\sqrt{\sum_{j=1}^p Cov(y, X_j)^2}} \sum_{j=1}^p Cov(y, X_j)^2 X_j$$

Peubah  $X_j$  dipilih yang berkorelasi tinggi  $\mathbf{Y}$  dan keragamannya cukup kuat. Seberapa penting peubah  $X_j$  dalam pembentukan komponen utama pertama  $t_1$  dilihat dengan membentuk model regresi linier sederhana  $\mathbf{Y}$  terhadap  $X_j$  yang dirumuskan sebagai:

$$Y = a_{1j}X_j + \epsilon_1$$

Nilai  $a_{1j}$  menunjukkan seberapa besar peran  $X_j$  pada pembentukan komponen pertama dan jika  $a_{1j}$  tidak signifikan maka peran  $X_j$  dapat diabaikan.

## 2. Metode Penelitian

Pembentukan estimasi model data Tingkat Kemiskinan di Indonesia tahun 2015 sebagai peubah respon  $\mathbf{Y}$  dengan lima peubah bebas yaitu  $X_1$ (AMH),  $X_2$ (rata-rata lama sekolah),  $X_3$  (APS/ angka partisipasi sekolah),  $X_4$ (pendapatan perkapita), dan  $X_5$  (pengangguran terbuka). Estimasi model dilakukan dengan menggunakan metode

PLSR, dan sebagai pembanding digunakan juga metode kuadrat terkecil (MKT). Data yang digunakan dalam estimasi model adalah data mengenai angka melek huruf (AMH), rata-rata lama sekolah, angka partisipasi sekolah, pendapatan perkapita, dan tingkat pengangguran terbuka terhadap tingkat kemiskinan yang dilihat dari persentase penduduk miskin di Indonesia tahun 2015.

### 3. Hasil dan Pembahasan

Data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi Bali untuk peubah-peubah penelitian dapat diringkas dalam Tabel 1.

Tabel 1. Statistika Deskriptif Data Penelitian

Peubah	Rataan	StDev	Minimum	Kuartil 1	Median	Kuartil 3	Maksimum
Y	12,20	6,49	3,72	6,87	10,39	16,14	31,35
$X_1$	94,63	4,80	75,92	91,70	96,85	97,70	99,56
$X_2$	8,30	0,91	6,05	7,58	8,27	9,06	10,60
$X_3$	21,82	6,61	9,46	17,46	20,25	24,11	45,86
$X_4$	40873	34906	12379	24332	30063	37177	157168
$X_5$	5,34	2,22	1,83	3,73	4,76	6,62	10,12

Sumber: Data diolah (2016)

Korelasi linear antar peubah penelitian diuraikan dalam Tabel 2.

Tabel 2. Korelasi Linier Antar Peubah Penelitian

Peubah	Y	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
Y	1	-0,570*	0,332*	0,239	-0,314	-0,113
		0,001	0,024	0,180	0,075	0,531
$X_1$		1	0,700*	0,007	0,275	0,433*
			0,000	0,970	0,122	0,012
$X_2$			1	0,321	0,547*	0,481*
				0,069	0,001	0,005
$X_3$				1	-0,105	0,131
					0,562	0,467
$X_4$					1	0,286
						0,106
$X_5$						1

Cell Contents: Pearson correlation  
P-Value

Matriks korelasi memperlihatkan bahwa hanya peubah  $X_1$  dan  $X_2$  yang signifikan korelasinya terhadap peubah respon Y pada taraf nyata 5%. Korelasi antar peubah bebas yang signifikan diantaranya  $X_1$  dan  $X_2$ ,  $X_1$  dan  $X_5$ ,  $X_2$  dan  $X_4$ , serta  $X_2$  dan  $X_5$ . Selanjutnya diuraikan proses estimasi model regresi, hasil uji serentak dan uji parsial dengan MKT ditampilkan dalam analisis ragam regresi pada Tabel 3. Tingkat

signifikansi dari suatu analisis untuk kasus bidang sosial umumnya memiliki nilai toleransi yang lebih tinggi dari penelitian bidang sains. Untuk kasus sosial tingkat kemiskinan di Indonesia, masih dimungkinkan untuk menetapkan taraf nyata 10%.

Tabel 3. Analisis Ragam Estimasi Model dengan MKT

Sumber Keragaman	Db	JK	KT	F-hit	P-Value
Regresi	5	576,46	115,291	4,03	0,007**
$X_1$	1	156,75	156,749	5,48	0,027**
$X_2$	1	3,14	3,138	0,11	0,743
$X_3$	1	48,81	48,811	1,71	0,203
$X_4$	1	13,72	13,721	0,48	0,495
$X_5$	1	31,66	31,661	1,11	0,302
Galat/Error	27	772,60	28,615		
Total	32	1349,05			
Model Summary					
S	R-sq	R-sq(adj)	R-sq(pred)		
5,34927	42,73%	32,13%	8,49%		
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	80,6	21,4	3,77	0,001	
$X_1$	-0,733	0,313	-2,34	0,027	2,53
$X_2$	-0,70	2,12	-0,33	0,743	4,16
$X_3$	0,232	0,178	1,31	0,203	1,55
$X_4$	-0,000025	0,000037	-0,69	0,495	1,84
$X_5$	0,520	0,494	1,05	0,302	1,34

Uji parsial memperlihatkan bahwa hanya peubah  $X_1$  yang signifikan masuk dalam model. Peubah  $X_2$  tidak masuk dalam model meskipun berkorelasi signifikan dengan peubah respon Y. Hal ini disebabkan karena terjadi multikolinearitas antara  $X_1$  dan  $X_2$  sehingga pengaruh  $X_2$  sudah diwakili oleh  $X_1$ . Walaupun antara peubah  $X_1$  berkorelasi secara signifikan dengan peubah  $X_2$ , namun berdasarkan nilai VIF tidak terjadi multikolinearitas yang kuat. Multikolinearitas dikatakan kuat apabila VIF lebih besar dari 5 [1].

Selanjutnya ditentukan estimasi model hanya dengan peubah  $X_1$ , diuraikan dalam analisis ragam regresi pada Tabel 4.

Tabel 4. Analisis Ragam Regresi MKT dengan Peubah Bebas  $X_1$ 

Sumber Keragaman	Db	JK	KT	F-Value	P-Value
Regresi	1	438,74	438,74	14,94	0,001
$X_1$	1	438,74	438,74	14,94	0,001
Galat/Error	31	910,31	29,36		
Lack-of-Fit	29	833,22	28,73	0,75	0,723
Pure Error	2	77,10	38,55		
Total	32	1349,05			

Model Summary			
S	R-sq	R-sq(adj)	R-sq(pred)
5,41895	32,52%	30,35%	21,18%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	85,1	18,9	4,51	0,000	
$X_1$	-0,771	0,199	-3,87	0,001	1,00

Regression Equation

$$Y = 85,1 - 0,771 X_1$$

Estimasi model dengan menggunakan MKT diperoleh:

$$\hat{Y} = 85,1 - 0,771 X_1$$

dengan koefisien determinasi  $R^2 = 32,52\%$ .

Model ini mengindikasikan bahwa keragaman data persentase jumlah penduduk miskin di Indonesia yang mampu dijelaskan oleh peubah angka melek huruf (AMH) sebesar 32,52%. Koefisien bertanda negatif menunjukkan bahwa semakin tinggi AMH yang mengindikasikan semakin tinggi tingkat pendidikan akan menurunkan persentase penduduk miskin di Indonesia.

Selanjutnya dilihat kinerja PLSR pada data dengan multikolinearitas yang tidak kuat, dengan bantuan software Minitab 17 metode *cross-validation: Leave-one-out* diperoleh model estimasi yang disajikan dalam analisis ragam pada Tabel 5.

Tabel 5. Analisis Ragam Estimasi Model PLSR

Sumber Keragaman	Db	JK	KT	F	P
Regresi	1	448,90	448,897	15,46	0,000
Residual Error	31	900,16	29,037		
Total	32	1349,05			

Model Selection and Validation Y						
Components	X Variance	Error	R-Sq	PRESS	R-Sq (pred)	
1	0,441302	900,155	0,332750	1124,79	0,166233	
2		786,392	0,417078	1155,07	0,143789	
3		774,645	0,425786	1183,05	0,123054	
4		773,015	0,426994	1214,07	0,100055	
5		772,595	0,427305	1234,46	0,084941	

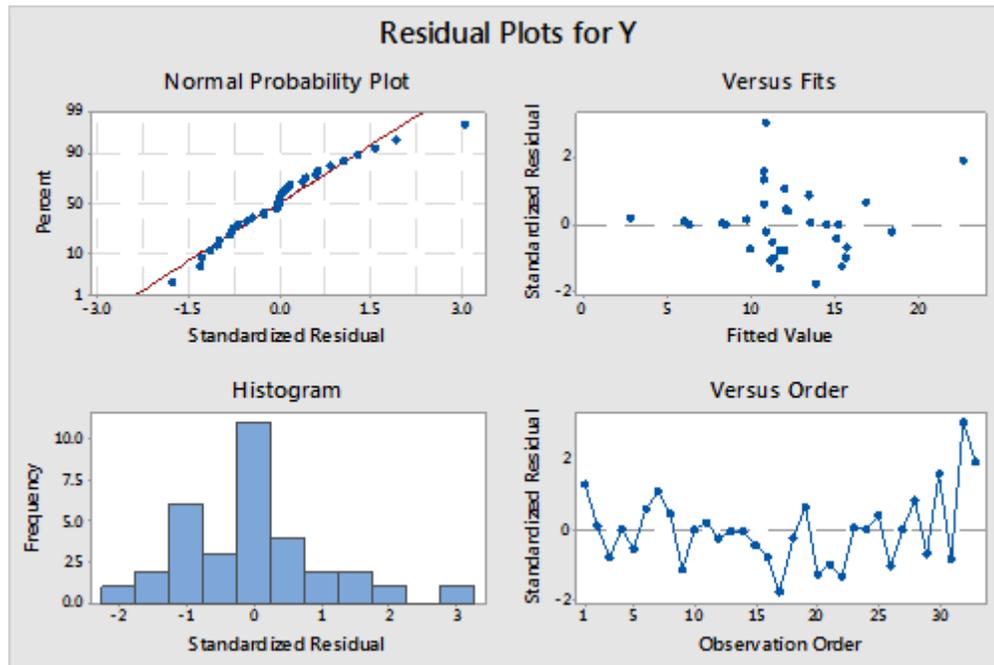
Coefficients		Y	
		Y	standardized
Constant		61,1363	0,000000
$X_1$		-0,39640	-0,293358
$X_2$		-1,43550	-0,201430
$X_3$		0,12070	0,122975
$X_4$		-0,00001	-0,161355
$X_5$		-0,17020	-0,058125

Estimasi model PLSR persentase penduduk miskin dengan lima peubah bebas yaitu:

$$\hat{Y} = 61.1363 - 0.3964 X_1 - 1.4355X_2 + 0.1207X_3 - 0.00001X_4 - 0.1702X_5$$

dengan koefisien determinasi  $R^2$  sebesar 33,275%,  $X_1$ :AMH,  $X_2$ : rataan lama sekolah,  $X_3$ : APS (angka partisipasi sekolah),  $X_4$ :pendapatan perkapita, dan  $X_5$ : pengangguran terbuka.

Hasil ini memperlihatkan bahwa satu komponen yang merupakan kombinasi linier dari peubah bebas mampu menerangkan keragaman data sebenarnya sebesar 33,275%. Hasil estimasi model dan nilai koefisien determinasi memperlihatkan pada data yang tidak mengandung multikolinearitas kuat dan jumlah peubah bebas jauh lebih sedikit dari jumlah data pengamatan, estimasi model dengan MKT dan PLSR memiliki koefisien determinasi yang tidak jauh berbeda. Kecilnya nilai koefisien determinasi dari estimasi model yang diperoleh disebabkan karena ada peubah-peubah bebas yang berpengaruh terhadap persentase penduduk miskin tetapi tidak dilibatkan dalam pembentukan model. Diagnostik model PLSR dengan melihat plot peluang normal galat, keacakan galat, plot galat dengan nilai dugaan seperti dalam Gambar 1.



Gambar 1. Plot Diagnostik Galat Estimasi Model

Gambar 1 menunjukkan bahwa kenormalan dan keacakan galat dari estimasi model PLSR telah terpenuhi.

#### 4. Kesimpulan

Pada data yang tidak mengalami kondisi buruk (*ill condition*), jumlah peubah bebas lebih sedikit dari jumlah data amatan, estimasi model regresi yang diperoleh dengan metode PLSR dan MKT menghasilkan koefisien determinasi yang hampir sama. Estimasi model data Tingkat Kemiskinan di Indonesia tahun 2015 yang diukur dengan persentase jumlah penduduk miskin sebagai peubah respon Y dengan lima peubah bebas yaitu  $X_1$ (AMH),  $X_2$ (rata-rata lama sekolah),  $X_3$  (APS/ angka partisipasi sekolah),  $X_4$ (pendapatan perkapita), dan  $X_5$  (pengangguran terbuka) dengan MKT diperoleh:

$$\hat{Y} = 85,1 - 0,771 X_1$$

dengan koefisien determinasi  $R^2 = 32,52\%$ .

Estimasi model PLSR diperoleh:

$$\hat{Y} = 61.1363 - 0.3964 X_1 - 1.4355X_2 + 0.1207X_3 - 0.00001X_4 - 0.1702X_5$$

dengan koefisien determinasi  $R^2$  sebesar 33,275%.

#### Daftar Pustaka

- [1] Montgomery, D.C. and Peck, E.A.(2007) '*Introduction to Linear Regression Analysis*', 3<sup>rd</sup> Ed., New York :John Wiley & Sons, Inc.
- [2] Abdi, H. (2006) '*Partial Least Squares Regression (PLSR)*', University of Texas
- [3] Esposito, V., Wynne, Henseler, J., and Wang, Huiwen (2010) '*Handbook of Partial least Squares*', Germany: Springer
- [4] Suryawati, C. (2005) '*Memahami Kemiskinan Secara Multidimensional*', [http://www.jmpk-online.net/Volume\\_8/Vol\\_08\\_No\\_03\\_2005.pdf](http://www.jmpk-online.net/Volume_8/Vol_08_No_03_2005.pdf). Diakses 3 Maret 2016
- [5] Sukirno, S. (1999) '*Makroekonomi Modern*', Jakarta: Penerbit Raja Grafindo Persada.
- [6] Forsman, A., Andersson, C., and Grimvall, A. (2003) 'Estimation of the impact of short-term fluctuations in input on temporally aggregated outputs of process-oriented models' *Journal of Hydroinformatics*, 5(3), 169-180.
- [7] Mevik, H. (2007) 'The PLS Package: Principal Component and Partial Least Square Regression in R' *Journal of Statistical Software*, vol 18, 1-24.
- [8] Somnath, D and Susmita, D. (2007) 'Predicting Patient Survival from Microarray Data by Accelerated failure Time Modeling using partial Least Squares and LASSO'. *Journal of Biometrics*, vol 63, 259-271. United State of America.
- [9] Subekti, R. dan Rosita, K. (2014) 'Kajian Partial Least Squares (Studi Kasus: Regresi Cox-PLS)', *Journal J.Sains Dasar*, vol 3 No. 1. Hal. 61-68.
- [10] Hudaya, D. (2009) 'Faktor-Faktor yang Mempengaruhi Tingkat Kemiskinan di Indonesia' *Skripsi*, Institut Pertanian Bogor. Bogor.