

ANALISIS KOMPONEN UTAMA UNTUK MENGATASI MULTIKOLINEARITAS PADA FAKTOR-FAKTOR YANG MEMPENGARUHI TINGKAT PENGANGGURAN TERBUKA DI PROVINSI NUSA TENGGARA TIMUR

Elisabeth Maubanu¹, Kartiko²

^{1,2}Jurusan Statistika, FST, Institut Sains & Teknologi AKPRIND Yogyakarta

E-mail:maubanu.elisabeth@gmail.com

Abstract. Regression analysis is used to investigate factors that affect to a particular variable. One important assumption that must be met is the absence of multicollinearity among independent variables. Multicollinearity results in regression coefficient variance not minimal, so regression model is unstable. Multicollinearity can be handled using principal component analysis method. This study using open unemployment rate data in East Nusa Tenggara province as well as factors that are suspected influence; Human Development Index, population density, population growth rate, growth rate of ADHK GRDP, Minimum Wage Regional, literacy rate, average school length and ratio of school (senior high school) per 10,000 population. In the analysis using OLS method, multicollinearity occurs on independent factors. Based on the results of analysis using the principal component analysis, there is no longer occur of multicollinearity on the regression estimator model. From 8 independent variables obtained 4 new variables that can explain 88.1% variance of the 8 variables. MSE of principal component regression model estimator is less than OLS method.

Keywords: Multicollinearity, Main Component Analysis, Open Unemployment Rate

Abstrak. Analisis regresi digunakan untuk menyelidiki faktor-faktor yang mempengaruhi suatu variabel tertentu. Salah satu asumsi penting yang harus dipenuhi adalah tidak adanya multikolinearitas antar variabel independen. Multikolinearitas berakibat pada tidak minimalnya variansi koefisien regresi, sehingga model regresi tidak stabil. Multikolinearitas dapat ditangani menggunakan metode analisis komponen utama. Penelitian ini menggunakan data tingkat pengangguran terbuka di provinsi Nusa Tenggara Timur serta faktor-faktor yang diduga mempengaruhi; Indeks Pembangunan Manusia, kepadatan penduduk, laju pertumbuhan penduduk, laju pertumbuhan PDRB ADHK, Upah Minimum Regional, angka melek huruf, rata-rata lama sekolah dan rasio jumlah sekolah (SMA/MA/SMK) per 10.000 jiwa penduduk. Dalam analisis menggunakan metode OLS, terjadi multikolinearitas pada faktor-faktor independen. Berdasarkan hasil analisis menggunakan analisis komponen utama, tidak lagi terjadi multikolinearitas pada penduga model regresi. Dari 8 buah variabel independen diperoleh 4 buah variabel baru yang mampu menjelaskan 88,1 % variansi dari ke-8 variabel tersebut. MSE dari penduga model regresi komponen utama lebih kecil dibandingkan menggunakan metode OLS.

Kata kunci: Multikolinearitas, Analisis Komponen Utama, Tingkat Pengangguran

1. Pendahuluan

Salah satu ilmu statistik yang seringkali digunakan adalah analisis regresi linear sederhana maupun berganda. Analisis regresi digunakan untuk menyelidiki faktor-faktor yang mempengaruhi suatu variabel tertentu. Metode estimasi regresi berganda yang sering digunakan adalah metode *Ordinary Least Squares (OLS)*. Salah satu kendala yang sering ditemui adalah tidak terpenuhinya asumsi klasik dalam sebuah pembuatan model regresi. Jika salah satu tidak terpenuhi maka hasil estimasi yang diperoleh pun tentu tidak menjelaskan secara signifikan dan sangat diragukan. Salah satu asumsi pada regresi OLS yang menarik untuk dibahas adalah masalah multikolinearitas. Multikolinearitas menyatakan adanya hubungan linear antar variabel independen. Suatu model regresi yang baik tidak diperkenankan adanya multikolinearitas. Jika pada saat melakukan estimasi menggunakan metode OLS terdapat multikolinearitas, maka hasil estimasi yang diperoleh bersifat tidak bias tetapi mempunyai variansi yang tidak minimum^[7]. Dalam menangani masalah multikolinearitas ada banyak metode.

Dalam penelitian ini pembahasan untuk mengatasi masalah multikolinearitas menggunakan analisis komponen utama (*PCA*). Analisis komponen utama adalah suatu teknik mereduksi data multivariat yang mengubah suatu matriks data yang banyak menjadi suatu set kombinasi linier yang lebih sedikit akan tetapi menyerap sebagian besar jumlah varian^[3]. Sedangkan studi kasus yang diangkat dalam penelitian ini adalah tingkat pengangguran di provinsi NTT dengan faktor-faktor yang diduga mempengaruhinya.

Provinsi NTT merupakan salah satu provinsi di Indonesia yang tergolong daerah tertinggal dengan memiliki angka laju pertumbuhan ekonomi rendah yaitu 5,15. Salah satu hal yang menjadi penyebabnya tentu adalah tingkat pengangguran yang tinggi menyebabkan tingkat perekonomian menjadi rendah. Sehingga dalam upaya membangun provinsi NTT perlu adanya penelitian-penelitian mengenai faktor-faktor yang mempengaruhi tingkat pengangguran di daerah agar setiap kebijakan yang diambil lebih efisien dan mampu meningkatkan perekonomian. Maka dalam penelitian ini akan dibahas faktor-faktor yang diduga mempengaruhi tingkat pengangguran di NTT dan menghilangkan masalah multikolinearitas yang terdapat pada faktor-faktor tersebut. Hingga memperoleh suatu penduga model regresi yang lebih baik dan bebas dari multikolinearitas.

2. Metode Penelitian

Adapun alur penelitian ini secara keseluruhan adalah sebagai berikut:

- a. Pengumpulan Data dan Penentuan Variabel
Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari web BPS Provinsi NTT dan Data Referensi Kementerian Pendidikan dan Kebudayaan Indonesia pada tahun 2015. Variabel penelitian meliputi; Tingkat Pengangguran Terbuka sebagai variabel dependen (Y), Indeks Pembangunan Manusia (X_1), Kepadatan Penduduk (X_2), Laju Pertumbuhan Penduduk (X_3), Laju Pertumbuhan Produk Domestik Regional Bruto (X_4) (Atas Dasar Harga Konstan), Upah Minimum Regional (X_5), Angka Melek Huruf (X_6), Rata-rata Lama Sekolah (X_7), dan Rasio Jumlah Sekolah (X_8), merupakan variabel-variabel independen.
- b. Melakukan analisis deskriptif
Analisis deskriptif digunakan untuk menggambarkan karakteristik data berupa ringkasan numerik yang meliputi nilai rata-rata, nilai minimum dan maksimum^[3], dari data tingkat pengangguran dan faktor-faktor yang mempengaruhi di Provinsi Nusa Tenggara Timur pada tahun 2015.
- c. Melakukan analisis regresi berganda menggunakan metode *Ordinary Least Square*
Estimasi parameter dengan metode *Ordinary Least Square (OLS)* merupakan langkah awal untuk melihat hubungan antara variabel tingkat pengangguran dengan faktor-faktor yang mempengaruhinya. Estimasi parameter dengan *OLS* dilakukan dengan meminimalkan jumlah kuadrat *error*. Bentuk umum persamaan regresi ganda dengan metode *OLS* adalah

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + \beta_8 X_{8i} + \varepsilon_i \quad (1)$$

dengan, $i=1,2,3..n$

Dalam analisis regresi berganda *OLS* terdapat beberapa asumsi yang harus dipenuhi yang meliputi uji normalitas, uji multikolinearitas, uji heteroskedastisitas dan uji autokorelasi.

- d. Mengidentifikasi multikolinearitas pada faktor-faktor independen
Metode identifikasi gejala multikolinearitas yang digunakan dalam penelitian ini adalah *variance inflation factor (VIF)*. Identifikasi ini dilakukan pada faktor-faktor yang mempengaruhi tingkat pengangguran di Provinsi Nusa Tenggara Timur pada tahun 2015. Formulasi atau rumus yang digunakan untuk menghitung *VIF* :

$$VIF = \frac{1}{1-R^2} \tag{2}$$

Apabila nilai VIF dari faktor-faktor independen ada yang lebih besar dari 10 maka terdapat multikolinearitas pada model regresi. Jika diperoleh hasil bahwa terdapat multikolinearitas maka dapat dilihat keeratan hubungan antar variabel yang saling berkorelasi dengan melihat matriks korelasi dan uji signifikansi korelasi Pearson yang diperoleh signifikan atau tidak. Hipotesis dalam uji signifikan korelasi Pearson adalah:

$$H_0 : r_{ij} = 0 \text{ (tidak ada hubungan)}$$

$$H_1 : r_{ij} \neq 0 \text{ (terdapat hubungan)}$$

Statistik uji : *p-value* hasil output minitab

Dengan menggunakan taraf signifikansi berupa α , H_0 ditolak jika *p-value* < α . Apabila H_0

ditolak berarti bahwa antara variabel *i* dan *j* terdapat hubungan linear.

e. Penanganan multikolinearitas menggunakan analisis komponen utama

Analisis komponen utama (PCA) adalah suatu analisis untuk mereduksi data multivariat yang banyak menjadi suatu kombinasi linier dari variabel-variabel yang saling berkorelasi yang lebih sedikit, akan tetapi masih menyerap sebagian besar jumlah informasi. Misalkan dari *k* buah variabel independen asal yang dapat menerangkan total variansi sistem, namun seringkali variansi tersebut sebenarnya dapat diterangkan dengan sejumlah kecil komponen utama, misalkan oleh *p* buah variabel saja, dimana $p < k$. Dalam hal ini *p* buah komponen mengganti *k* buah variabel asal atau dikatakan dapat direduksi ke ukuran yang lebih kecil.

Komponen utama didefinisikan sebagai kombinasi linear terbobot dari variabel asal dengan jumlah *k* buah variabel asal adalah sebagai berikut:

$$\begin{aligned} w_1 &= a_{11} X_1 + a_{21} X_2 + a_{31} X_3 + \dots + a_{k1} X_k \\ w_2 &= a_{12} X_1 + a_{22} X_2 + a_{32} X_3 + \dots + a_{k2} X_k \\ w_3 &= a_{13} X_1 + a_{23} X_2 + a_{33} X_3 + \dots + a_{k3} X_k \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ w_k &= a_{1k} X_1 + a_{2k} X_2 + a_{3k} X_3 + \dots + a_{kk} X_k \end{aligned} \tag{2}$$

Apabila data dari variabel-variabel independen yang diamati tidak semuanya menggunakan satuan pengukuran yang sama maka variabel tersebut perlu distandarisasikan menggunakan rumus:

$$Z_{ji} = \frac{(x_{ji} - \bar{x}_j)}{s_k} \tag{3}$$

untuk $i=1,2,3,\dots,n$

$j=1,2,3,\dots,k$

k = jumlah variabel independen

\bar{x}_k : rata-rata sampel x_k

s_k : standar deviasi dari x_k

Pembentukan komponen utama dapat menggunakan matriks *S* maupun matriks korelasi (**R**). Perhitungan menggunakan matriks **R** apabila data memiliki satuan ukuran yang berbeda-beda dengan syarat data yang digunakan berupa bentuk baku. Untuk menghitung koefisien komponen utama digunakan proses iterasi dari matriks **R** dengan penggandaan matriks **R**. Selanjutnya untuk memperoleh vektor awal (\mathbf{a}'_k) ditentukan vektor awal dengan mempertimbangkan matriks **R** yaitu \mathbf{a}'_0 yang disesuaikan berdasarkan matriks **R**.

Untuk mengetahui elemen iterasi dengan menghitung Hasil iterasi $i = (\mathbf{a}'_0 \mathbf{R}^m)$ /elemen terbesar, dengan $n=2,4,6,8,\dots$.

Proses iterasi berhenti jika diperoleh hasil iterasi yang diperoleh berturut-turut mendapatkan nilai yang sama.

Dari beberapa kombinasi linear yang terbentuk pada analisis komponen utama, kriteria yang diinginkan adalah variansi komponen utama pertama bernilai maksimum. Oleh sebab itu perlu menentukan vektor koefisien pembobot komponen utama pertama (w_i) yang memaksimumkan variansi komponen utama pertama (Gaspersz:1995) dengan kendala:

$$\mathbf{a}'_k \mathbf{a}_k = 1 \quad (4)$$

Dengan kendala tersebut maka misalkan diperoleh hasil iterasi yang sama sebagai berikut :

$$[\mathbf{a}_{01k} \quad \mathbf{a}_{02k} \quad \mathbf{a}_{03k} \quad \mathbf{a}_{04k} \quad \mathbf{a}_{05k}]$$

Maka dengan kendala $\mathbf{a}'_k \mathbf{a}_k = 1$:

$$\begin{aligned} \mathbf{a}_{1k} &= \frac{\mathbf{a}_{01k}}{\sqrt{\mathbf{a}_{01k}^2 + \mathbf{a}_{02k}^2 + \mathbf{a}_{03k}^2 + \mathbf{a}_{04k}^2 + \mathbf{a}_{05k}^2}} \\ \mathbf{a}_{2k} &= \frac{\mathbf{a}_{02k}}{\sqrt{\mathbf{a}_{01k}^2 + \mathbf{a}_{02k}^2 + \mathbf{a}_{03k}^2 + \mathbf{a}_{04k}^2 + \mathbf{a}_{05k}^2}} \\ &\vdots \\ \mathbf{a}_{5k} &= \frac{\mathbf{a}_{05k}}{\sqrt{\mathbf{a}_{01k}^2 + \mathbf{a}_{02k}^2 + \mathbf{a}_{03k}^2 + \mathbf{a}_{04k}^2 + \mathbf{a}_{05k}^2}} \end{aligned}$$

Sehingga diperoleh $\mathbf{a}'_k = [\mathbf{a}_{1k} \quad \mathbf{a}_{2k} \quad \mathbf{a}_{3k} \quad \mathbf{a}_{4k} \quad \mathbf{a}_{5k}]$

Maka bentuk komponen utama menjadi:

$$w_k = \mathbf{a}_{1k} z_1 + \mathbf{a}_{2k} z_2 + \mathbf{a}_{3k} z_3 \dots + \mathbf{a}_{kk} z_k \quad (5)$$

Selanjutnya untuk menghitung nilai eigen menggunakan persamaan:

$$(\mathbf{R} - \lambda \mathbf{I}) \mathbf{A} = \mathbf{0},$$

dengan λ adalah akar karakteristik (nilai eigen) dan \mathbf{I} adalah matriks identitas. Apabila nilai eigen adalah $\lambda_1, \lambda_2, \dots, \lambda_k$, dan $\lambda_1 > \lambda_2 > \lambda_3 \dots > \lambda_k$.

Pemilihan komponen utama membutuhkan beberapa kriteria antara lain proporsi kumulatif variansi data asal yang dijelaskan oleh k komponen utama minimal 80% ,atau berdasarkan scree plot ditentukan dengan melihat letak terjadinya belokan dengan menghapus komponen utama yang memiliki nilai eigen lebih kecil 1 yang membentuk pola garis lurus, atau komponen yang dipilih adalah komponen utama yang memiliki nilai eigen terbesar atau lebih besar dari 1.

f. Melakukan analisis regresi komponen utama

Tahap analisis regresi komponen utama dilakukan setelah memperoleh komponen utama yang layak kemudian diregresikan dengan variabel Y dengan langkah yang sama dngan metode OLS. Apabila pada tahap awal pembentukan komponen utama, data sudah dilakukan standarisasi maka setelah memperoleh bentuk persamaan regresi komponen utama, nilai Z_k harus ditransformasikan kembali menjadi X_k .

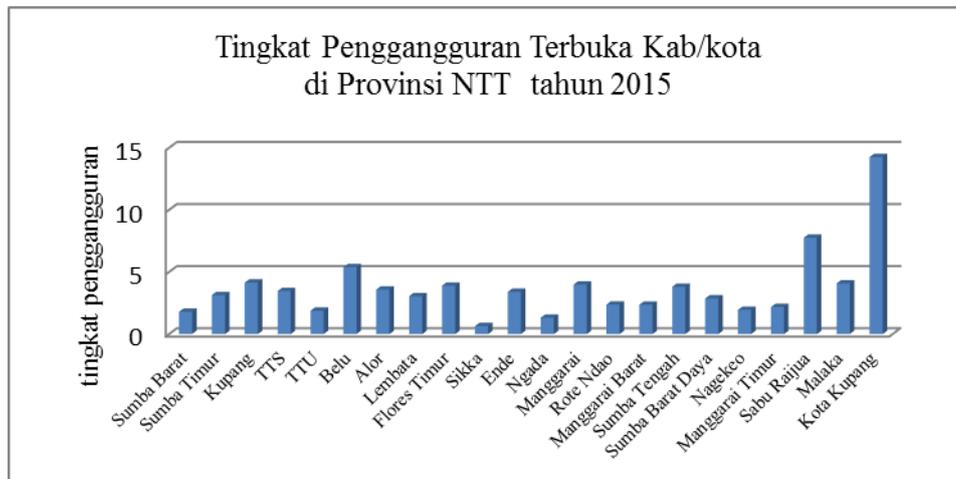
g. Menginterpretasikan hasil analisis regresi komponen utama

Dari hasil regresi komponen utama yang diperoleh maka dapat diinterpretasikan hasil penduga model regresi tingkat penggangguran terbuka di NTT oleh fakto- faktor yang mempengaruhinya.

3. Hasil Dan Pembahasan

Statistik Deskriptif

Data tingkat penggangguran terbuka Provinsi NTT dan faktor-faktor yang mempengaruhinya pada tahun 2015 yang masing-masing berjumlah 22 data yang diambil berdasarkan Kabupaten/Kota yang ada ditunjukkan dengan diagram batang pada Gambar 1.



Gambar 1. Diagram Batang Tingkat Penggauran Terbuka NTT tahun 2015

Sedangkan untuk faktor-faktor yang diduga mempengaruhinya ada 8 buah variabel. Deskriptifnya dapat dilihat pada Tabel 1.

Tabel 1. Deskriptif statistik

Variabel	Mean	Standarr Deviasi	Minimum	Median	Maximum
X_1	61,285 %	4,623	53,28 %	60,915	77,95
X_2	224 jiwa/km ²	495	35 jiwa/km ²	116	2432 jiwa/km ²
X_3	1,726 %	0,863	0,53%	1,645	3,99 %
X_4	4,937 %	0,575	4,14%	4,975	6,95%
X_5	Rp 1.464.871	142366	1361507	1401944	1813603
X_6	92,413 %	4485	83,77%	93,315	98,88%
X_7	6,903 tahun	1,197	5,12 tahun	6,81	11,43 tahun
X_8	1,576	0,449	1,022	1,531	2,701

Estimasi parameter dengan metode OLS

Dengan menggunakan metode OLS diperoleh estimasi parameter model regresi untuk tingkat penggangguran terbuka kabupaten /kota di provinsi NTT pada tahun 2015, sehingga didapat penduga model regresi sebagai berikut:

$$\hat{y} = 12,1 - 0,320X_1 + 0,00524 X_2 + 0,028 X_3 + 1,57 X_4 + 1,54 \times 10^{-6} X_5 - 0,0184X_6 + 0,290 X_7 - 0,176 X_8 \tag{6}$$

Dari penduga model regresi ini memiliki nilai koefisien determinasi sebesar 0,76. Nilai koefisien determinasi tersebut menjelaskan bahwa bahwa variabel-variabel independen telah mampu menjelaskan 76% variabel dependen.

Deteksi Multikolinearitas

Pendeteksian multikolinearitas menggunakan metode VIF , apabila nilai VIF > 10 maka dapat disimpulkan bahwa terdapat multikolinearitas. Dengan bantuan *software* Minitab diperoleh hasil sebagai berikut:

Tabel 2. Nilai VIF variabel independen

Variabel	VIF
X_1	7,281
X_2	9,282
X_3	1,668
X_4	3,387
X_5	1,633
X_6	2,074
X_7	16,085
X_8	1,895

Nilai VIF pada variabel $X_7 = 16,085$ maka disimpulkan bahwa terdapat multikolinearitas. Untuk lebih jelasnya dapat dilihat dari matriks korelasi Pearson:

		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		1	0,792	-0,011	0,629	0,043	0,450	0,885	0,042
X_2		0,792	1	0,289	0,790	-0,150	0,286	0,836	0,003
X_3		-	0,289	1	0,375	-0,354	-0,031	0,095	-0,129
X_4		0,629	0,790	0,375	1	-0,000	0,249	0,641	-0,046
X_5		0,043	-0,150	-0,354	-0,000	1	-0,103	0,058	0,209
X_6		0,450	0,286	-0,031	0,249	-0,103	1	0,552	0,212
X_7		0,885	0,836	0,095	0,641	0,058	0,552	1	0,297
X_8		0,042	0,003	-0,129	-0,046	0,209	0,212	0,297	1

Dari matriks korelasi tersebut diketahui bahwa terdapat beberapa variabel yang memiliki r_{ij} mendekati satu ($> 0,75$) yang berarti bahwa kedua variabel tersebut memiliki hubungan linear yang kuat (berkorelasi). Misalnya antara X_1 dan X_2 memiliki nilai koefisien korelasi 0,792 berarti antar variabel X_1 dan X_2 memiliki korelasi satu sama lain, demikian juga selanjutnya untuk variabel-variabel yang lain.

Analisis komponen utama

Berdasarkan uji asumsi multikolinearitas yang telah dibahas sebelumnya, diperoleh bahwa data yang digunakan mengandung multikolinearitas. Oleh sebab itu untuk mengatasi multikolinearitas tersebut, maka dilakukan analisis komponen utama. Data yang digunakan memiliki satuan ukuran yang berbeda-beda sehingga langkah pertama adalah mengubahnya dalam bentuk baku.

Misalkan perhitungan *z-score* untuk variabel X_1 adalah sebagai berikut:

$$\text{Rata-rata : } \bar{x}_1 = \frac{\sum x_1}{n} = \frac{1348,3}{22} = 61,285$$

$$\begin{aligned} \text{Standar deviasi (s}_1) &= \sqrt{\frac{\sum (x_{1i} - \bar{x}_1)^2}{n-1}} \\ &= \sqrt{\frac{(448,74)}{21}} \\ &= 4,623 \end{aligned}$$

$$Z_{11} = \frac{(X_{11} - \bar{x}_1)}{s_1} = \frac{(61,36 - 61,285)}{4,623} = 0,0163$$

$$Z_{12} = \frac{(X_{12} - \bar{x}_1)}{s_1} = \frac{(62,54 - 61,285)}{4,623} = 0,2716$$

⋮

$$Z_{1;22} = \frac{(X_{122} - \bar{x}_1)}{s_1} = \frac{(77,95 - 61,285)}{4,623} = 3,6049$$

Dengan menggunakan data *z-score* dari masing-masing variabel, dilakukan analisis komponen utama, dengan menggunakan bantuan *software* Minitab terbentuk 8 buah komponen utama. Komponen utama yang diperoleh disajikan dalam Tabel 3.

Tabel 3 Komponen utama yang terbentuk

	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8
Z_1	0,473	0,119	-0,139	-0,250	-0,164	0,357	0,638	0,345
Z_2	0,485	-0,144	-0,143	0,022	-0,282	-0,007	-0,666	0,448
Z_3	0,130	-0,580	0,129	0,554	0,379	0,409	0,090	0,067
Z_4	0,431	-0,185	-0,278	0,177	0,172	-0,745	0,260	-0,140
Z_5	-0,032	0,529	-0,598	0,246	0,489	0,176	-0,139	0,107
Z_6	0,289	0,214	0,574	-0,323	0,620	-0,107	-0,124	0,160
Z_7	0,498	0,176	0,043	0,024	-0,102	0,300	-0,161	-0,770
Z_8	0,074	0,488	0,421	0,659	-0,293	-0,137	0,116	0,165

Dengan nilai eigen dan proporsi dari masing-masing komponen sebagai berikut:

Tabel 4 Nilai proporsi kumulatif dan nilai eigen

	W_1							
Cumulative proporsi	0,449	0,650	0,775	0,881	0,947	0,981	0,995	1
Nilai eigen	3,592	1,609	0,998	0,847	0,528	0,271	0,116	0,038

Dari hasil ini digunakan kriteria pemilihan komponen utama nilai proporsi variansinya >80 %. Dengan demikian komponen utama yang terpilih adalah komponen pertama sama dengan ke-4 sehingga ada 4 komponen utama.

Langkah selanjutnya adalah menghitung score dari keempat komponen utama yang terpilih. Dengan berdasarkan Tabel 3. diperoleh persamaan untuk menghitung *score* komponen utama sebagai berikut:

$$\begin{aligned} W_1 &= 0,473 Z_1 + 0,485 Z_2 + 0,130 Z_3 + 0,431 Z_4 - 0,032 Z_5 + 0,289 Z_6 + 0,498 Z_7 + 0,074 Z_8 \\ W_2 &= 0,119 Z_1 - 0,144 Z_2 - 0,580 Z_3 - 0,185 Z_4 + 0,529 Z_5 + 0,214 Z_6 + 0,176 Z_7 + 0,488 Z_8 \\ W_3 &= -0,139 Z_1 - 0,143 Z_2 + 0,129 Z_3 - 0,278 Z_4 - 0,598 Z_5 + 0,574 Z_6 + 0,043 Z_7 + 0,421 Z_8 \\ W_4 &= -0,250 Z_1 + 0,022 Z_2 + 0,554 Z_3 + 0,177 Z_4 + 0,246 Z_5 - 0,323 Z_6 + 0,024 Z_7 + 0,659 Z_8 \end{aligned} \tag{7}$$

Maka diperoleh score komponen utama yang diregresikan dengan variabel independen. Selanjutnya adalah melakukan analisis regresi komponen utama dengan menggunakan data 4 buah *score* komponen utama dengan variabel *z-score* dari variabel *Y* seperti dibawah ini:

Tabel 5 *Score* komponen utama

No.	Z_y	W_1	W_2	W_3	W_4
1	-0,6802	-0,94079	-0,94430	-1,18546	-0,04891
2	-0,2000	-0,25317	-0,28797	-0,08821	-0,98972
3	0,1619	0,30828	-0,25795	1,28562	1,60307
4	-0,0818	-1,43048	0,38535	-0,92672	-0,82584
5	-0,6515	-0,71905	2,25202	-1,16189	0,84044
6	0,6063	0,08928	0,38814	-1,54316	0,64691
7	-0,0388	-0,34596	2,08243	2,09309	0,70806
8	-0,2323	0,46549	1,30950	-0,48071	0,96804
9	0,0723	-0,14386	1,34264	-0,64671	0,00875
10	-1,0923	-0,72453	0,19185	0,37765	-1,71642
11	-0,1069	0,81855	1,87063	-0,87295	-0,80538
12	-0,8522	0,80462	0,32071	0,99070	-0,99042
13	0,1046	0,15737	-0,66461	0,30472	-0,91178
14	-0,4760	-0,18248	-2,57026	0,41350	0,69569
15	-0,4795	-0,52180	-0,44067	1,44432	-0,32509
16	0,0401	-1,56835	-1,28263	-0,17451	-0,81100
17	-0,2968	-1,03220	-0,53705	-0,38039	1,03321
18	-0,6265	0,13690	0,15480	0,48944	-1,30253
19	-0,5440	-0,08695	0,53085	1,63876	0,70509
20	1,4556	-1,44329	-2,26036	-0,31524	0,58995
21	0,1368	-1,26533	-0,65425	-0,78361	0,84659
22	3,7813	7,87775	-0,92888	-0,47823	0,08128

Dengan menggunakan teknik analisis regresi berganda diperoleh penduga model regresi komponen utama sebagai berikut:

$$\hat{z}_y = -4,649 X 10^{-6} + 0,389 W_1 - 0,218 W_2 - 0,218 W_3 + 0,228 W_4 \quad (8)$$

Koefisien determinasi dari penduga model regresi komponen utama sebesar 0,6452, yang berarti bahwa variabel independen yang berupa komponen utama telah mampu menjelaskan 64,52% variabel dependen sedangkan sisanya 35,48 % dijelaskan oleh variabel lainya yang tidak dimasukan dalam penelitian. Akan tetapi dalam identifikasi multikolinearitas pada variabel-variabel baru ini tidak lagi mengandung multikolinearitas. Hal ini dibuktikan dengan nilai VIF dari masing-masing variabel bernilai 1. Jadi terbukti bahwa analisis komponen utama telah mampu menghilangkan masalah multikolinearitas yang sebelumnya terjadi pada variabel asal. Persamaan (8) ini bisa dikembalikan ke variabel asal dengan kebalikan menggunakan persamaan (7) dan (3) maka akan mendapatkan persamaan

$$\hat{Y} = -6,6506 + 0,0793X_1 + 0,0014X_2 + 0,8934X_3 + 1,4874X_4 + 1,1445 \times 10^{-6}X_5 - 0,0828 X_6 + 0,3529X_7 - 0,1197X_8 \quad (9)$$

4. Kesimpulan

Berdasarkan analisis dan pembahasan yang dilakukan, diperoleh bahwa tingkat penggangguran terbuka di provinsi NTT pada tahun 2015 memiliki rata-rata sebesar 3,698%. Presentase tertinggi tingkat penggangguran terbuka di provinsi NTT mencapai 14,25 % pada Kota Kupang sedangkan presentase terendah sebesar 0,65 % pada kabupaten Sikka. Faktor-faktor yang diduga mempengaruhi tingkat penggangguran di Provinsi NTT ada 8 faktor; Indeks Pembangunan Manusia (IPM), kepadatan penduduk, laju pertumbuhan penduduk, laju pertumbuhan PDRB ADHK, Upah Minimum Regional (UMR), angka melek huruf, rata-rata lama sekolah dan rasio jumlah sekolah (SMA/MA/SMK) per 10.000 jiwa penduduk.

Dengan menggunakan metode OLS diperoleh penduga model regresi untuk tingkat penggangguran terbuka di provinsi NTT sebagai berikut:

$$\hat{y} = 12,1 - 0,320X_1 + 0,00524 X_2 + 0,028 X_3 + 1,57 X_4 + 1,54 \times 10^{-6} X_5 - 0,0184X_6 + 0,290 X_7 - 0,176 X_8$$

Berdasarkan penduga model regresi ini, variabel-variabel independen telah mampu menjelaskan 76 % variabel dependen akan tetapi pada pegujian asumsi diperoleh kesimpulan adanya multikolinearitas.

Dari pembahasan terdapat multikolinearitas pada variabel-variabel independen yang dibuktikan dengan terdapat nilai VIF dari variabel yang lebih dari 10 yaitu variabel rata-rata lama sekolah. Untuk mengatasi multikolinearitas menggunakan analisis komponen utama. Dengan metode ini diperoleh 4 komponen utama yang layak digunakan untuk penduga model regresi komponen utama karena mempunyai kumulatif proporsi variansi sebesar 0,881 yang berarti bahwa sudah mampu menjelaskan 88,1 % variansi dari variabel asal. Penduga model regresi komponen utama yang diperoleh dari hasil regresi analisis komponen menggunakan 4 variabel baru tersebut adalah:

$$\hat{z}_y = -4,649 X 10^{-6} + 0,389 W_1 - 0,218 W_2 - 0,218 W_3 + 0,228 W_4$$

Model regresi komponen utama dianggap lebih baik dibandingkan dengan model regresi yang diperoleh menggunakan metode OLS karena tidak lagi mengandung multikolinearitas antar variabel independen dan memiliki nilai MSE lebih kecil yaitu 0,2742 dibandingkan nilai MSE dari metode OLS yang sebesar 1,1043.

Ucapan Terima Kasih

Puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa karena atas berkat dan rahmatNya penulis telah menyelesaikan penelitian ini dengan baik. Ucapan terima kasih juga penulis haturkan buat orang tua, saudara-saudara, dan sahabat-sahabat yang telah mendukung baik secara materi maupun doa. Ucapan terimakasih juga diberikan kepada bapak/ibu dosen jurusan Statistika IST AKPRIND Yogyakarta yang telah membimbing mulai dari penyusunan sampai penyelesaian penelitian ini.

Daftar Pustaka

- [1] Farid, 2008, *Analisis Tingkat Pengangguran Di Indonesia Tahun 1980-2007*, Fakultas Ekonomi Universitas Diponegoro, Semarang.
- [2] Gaspersz Vincent, 1995, *Teknik Analisis dalam Penelitian Percobaan 2*, Tarsito, Bandung.
- [3] Harinaldi, 2005, *Prinsip-Prinsip Statistik untuk Teknik dan Sains*, Erlangga, Jakarta.
- [4] Marcus dkk, 2012, *Analisis Regresi Komponen Utama untuk Mengatasi Multikolinearitas dalam Analisis Regresi Berganda*, Jurnal Berekeng Vol.6 No.1 Hal 31-40.
- [5] Morrison, 1990, *Multivariate Statistical Methods Third Edition*, McGraw-Hill, Singapore
- [6] Suryanto, 1998, *Metode Statistika Multivariat*, Dep. Pendidikan dan Kebudayaan, Jakarta.
- [7] Widarjono, 2014, *Analisis Multivariat Terapan*, UPP STIM YKPN, Yogyakarta.
- [8] Widiharini, 2001, *Penanganan Multikolinearitas (Kekolinearan Ganda) dengan Analisis Komponen Utama*, Jurnal Matematika dan Terapan Vol.4, No.2

- [9] _____, 2015, *Data Referensi Pendidikan Per Provinsi (Provinsi Nusa Tenggara Timur)*, Kementrian Pendidikan dan Kebudayaan Indonesia, diakses pada tanggal 06 April 2017 pada situs <http://referensi.data.kemdikbud.go.id/>
- [10] _____, 2015, *Provinsi Nusa Tenggara Timur Dalam Angka 2015*, BPS, Nusa Tenggara Timur, diakses pada tanggal 22 Maret 2017 pada situs <http://www.ntt.bps.go.id/>
- [11] _____, 2015, *Statistik Daerah 2015*, BPS, Nusa Tenggara Timur, diakses tanggal 17 Maret 2017 pada situs <http://www.ntt.bps.go.id/>

