

PEMODELAN INDEKS PEMBANGUNAN MANUSIA DI JAWA TENGAH DENGAN REGRESI KOMPONEN UTAMA *ROBUST*

Tsania Faizia¹, Alan Prahutama², Hasbi Yasin³

^{1,2,3}Departemen Statistika FSM Universitas Diponegoro

Alan.prahutama@gmail.com

ABSTRACT

Robust principal component regression is development of principal component regression that applies robust method at principal component analysis and principal component regression analysis. Robust principal component regression does not only overcome multicollinearity problems, but also overcomes outlier problems. The robust methods used in this research are Minimum Covariance Determinant (MCD) that is applied when doing principal component analysis and Least Trimmed Squares (LTS) that is applied when doing principal component regression analysis. The case study in this research is Human Development Index (HDI) in Central Java in 2017 which is influenced by labor force participation rates, school enrollment rates, percentage of poor population, population aged 15 years and over who are employed, health facilities, gross enrollment rates, and net enrollment rates. The model of HDI in Central Java in 2017 using robust principal component regression MCD-LTS provides the most effective result for handling multicollinearity and outliers with Adjusted R^2 value of 0.6913 and RSE value of 0.469.

Keywords: Robust Principal Component Regression, Multicollinearity, Outliers, Minimum Covariance Determinant (MCD), Least Trimmed Squares (LTS), Human Development Index (HDI).

1. PENDAHULUAN

Multikolinieritas merupakan salah satu masalah yang sering dihadapi dalam analisis regresi berganda. Multikolinieritas menyebabkan estimator mempunyai varian yang besar sehingga sulit mendapatkan estimasi yang tepat. Akibatnya, interval estimasi akan cenderung lebih besar dan nilai statistik uji t akan kecil sehingga menyebabkan variabel prediktor secara statistik tidak berpengaruh nyata (signifikan) terhadap variabel respon, meskipun nilai koefisien determinasi (R^2) masih relatif tinggi (Widarjono, 2005). Oleh karena itu, diperlukan penanganan terhadap multikolinieritas agar diperoleh model regresi yang stabil.

Salah satu metode untuk mengatasi multikolinieritas adalah regresi komponen utama (Montgomery *et al.*, 2012). Regresi komponen utama menangani multikolinieritas dengan dua tahap. Tahap pertama yaitu melakukan analisis komponen utama menggunakan *eigenvector* dari matriks kovarian atau matriks korelasi variabel prediktor. Tahap kedua yaitu meregresikan komponen-komponen utama terpilih dengan variabel respon menggunakan metode *Ordinary Least Squares* (OLS).

Menurut Filzmoser (1999), analisis komponen utama klasik tidak dapat bekerja dengan baik jika data yang dianalisis mengandung *outlier*. Hal ini dikarenakan vektor rata-rata dan matriks kovarian atau matriks korelasi sampel sangat sensitif terhadap *outlier* dan akan berpengaruh pada keakuratan hasil analisis komponen utama. Kemudian, menurut Srinadi (2014), metode *Ordinary Least Squares* (OLS) pada analisis regresi mempunyai kelemahan ketika *outlier* terdapat dalam data yang menyebabkan estimator dari parameter regresi bersifat bias. Oleh karena itu, regresi komponen utama dikembangkan menjadi

regresi komponen utama *robust*, yang menerapkan metode *robust* pada kedua tahap tersebut.

Metode *robust* yang digunakan dalam penelitian ini adalah metode *Minimum Covariance Determinant* (MCD) yang diterapkan pada saat melakukan analisis komponen utama dan metode *Least Trimmed Squares* (LTS) yang diterapkan pada saat melakukan analisis regresi komponen utama. Metode MCD dipilih karena Rousseeuw dan Driessen (1999) menyatakan bahwa metode tersebut lebih unggul dari metode lain baik dari segi efisiensi maupun kecepatan perhitungan statistik. Selain itu, metode tersebut sangat *robust* dalam penentuan lokasi dan sebaran multivariat. Sehingga, metode MCD sangat berguna dalam menghasilkan analisis yang lebih handal. Metode LTS dipilih karena menurut Rousseeuw dan Driessen (2006), metode LTS lebih efisien dan memiliki fungsi objektif yang lebih *smooth* (halus) sehingga kurang sensitif terhadap efek lokal. Selain itu, menurut Perihatini (2018), metode LTS merupakan salah satu metode penaksiran parameter model regresi yang *robust* terhadap kehadiran *outlier* yang memiliki nilai *breakdown point* yang tinggi dibandingkan metode yang lainnya.

Indeks Pembangunan Manusia (IPM) merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia. IPM dibangun dari tiga dimensi dasar yang digunakan sebagai ukuran kualitas hidup manusia. Dimensi tersebut mencakup kesehatan, pendidikan, dan ekonomi.

Adapun variabel-variabel yang dapat digunakan untuk mengukur masing-masing dimensi dasar IPM tersebut adalah sebagai berikut:

1. Untuk mengukur dimensi kesehatan dapat digunakan banyaknya sarana kesehatan (Putra dan Ratnasari, 2015).
2. Untuk mengukur dimensi pendidikan dapat digunakan angka partisipasi kasar dan angka partisipasi murni (Ekosiswoyo *et al.*, 2008).
3. Untuk mengukur dimensi ekonomi dapat digunakan presentase penduduk miskin, jumlah penduduk usia 15 tahun ke atas yang bekerja, dan tingkat partisipasi angkatan kerja (Putra dan Ratnasari, 2015).

Menurut Putra dan Ratnasari (2015), faktor-faktor di setiap dimensi dasar IPM akan cenderung memiliki hubungan yang kuat satu sama lainnya, karena antar faktor-faktor tersebut saling mempengaruhi. Sehingga dapat mengakibatkan adanya kasus multikolinieritas.

Dari uraian tersebut, tujuan utama dalam penelitian ini adalah melakukan penanganan multikolinieritas dan *outlier* pada pemodelan Indeks Pembangunan Manusia di Jawa Tengah menggunakan regresi komponen utama *robust* yang menerapkan metode MCD pada saat melakukan analisis komponen utama dan metode LTS pada saat melakukan analisis regresi komponen utama. Selain itu, penelitian ini juga bertujuan untuk mengukur efektivitas penerapan metode MCD dan LTS dalam regresi komponen utama *robust* untuk pemodelan Indeks Pembangunan Manusia di Jawa Tengah.

2. TINJAUAN PUSTAKA

2.1 Analisis Regresi Regresi Berganda

Menurut Montgomery *et al.* (2012), Model regresi berganda dinyatakan dalam bentuk persamaan:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad (1)$$

Jika ditulis dalam bentuk matriks, persamaan (1) dapat ditulis sebagai berikut:

$$y = X\beta + \varepsilon \quad (2)$$

dimana

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix}; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}; \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

2.1.1 Estimasi Parameter Model

Menurut Montgomery *et al.* (2012), estimasi parameter dapat diperoleh dengan menggunakan metode *Ordinary Least Squares* (OLS) yaitu dengan meminimumkan jumlah kuadrat *error*.

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

Untuk mendapatkan estimator OLS ($\hat{\boldsymbol{\beta}}$) yang meminimumkan $S(\boldsymbol{\beta})$ disyaratkan bahwa:

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0}$$

maka

$$\begin{aligned} \frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} &= \mathbf{0} \\ -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{0} \\ \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{y} \\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y}) \end{aligned}$$

Menurut Kutner *et al.* (2005), perbedaan unit satuan pada model regresi yang tidak distandarkan dapat menyebabkan koefisien regresi tidak bisa dibandingkan. Oleh karena itu, perlu dilakukan standarisasi menggunakan rumus sebagai berikut:

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \quad \text{dan} \quad X_{ji}^* = \frac{X_{ji} - \bar{X}_j}{S_j}$$

$$\text{dimana } S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} \quad \text{dan} \quad S_j = \sqrt{\frac{\sum_{i=1}^n (X_{ji} - \bar{X}_j)^2}{n-1}}$$

Sehingga diperoleh model regresi standar sebagai berikut:

$$Y_i^* = \beta_1^* X_{1i}^* + \beta_2^* X_{2i}^* + \cdots + \beta_k^* X_{ki}^* + \varepsilon_i^*$$

2.1.2 Uji Signifikansi dalam Regresi Berganda

Menurut Montgomery *et al.* (2012), dalam analisis regresi berganda ada beberapa uji signifikansi yang berguna untuk mengukur ketepatan model, antara lain sebagai berikut:

1. Uji Signifikansi Model Regresi

Uji ini dimaksudkan untuk menentukan apakah ada hubungan linier antara variabel respon Y dengan variabel prediktor X_1, X_2, \dots, X_k atau tidak. Adanya hubungan linier antara variabel respon dengan variabel prediktor menandakan bahwa model regresi yang terbentuk sesuai. Berikut langkah-langkahnya:

a. Hipotesis

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ (Model regresi tidak sesuai)

$H_1: \text{ada minimal satu } \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, k$ (Model regresi sesuai)

b. Statistik uji

$$F_0 = \frac{SSR/k}{SS_E/(n-k-1)} = \frac{MSR}{MSE}; \quad SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2; \quad SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

c. Kriteria uji

H_0 ditolak jika $F_0 > F_{\text{tabel}} = F(\alpha; k; n-k-1)$ atau $p\text{-value} < \alpha$.

2. Uji Signifikansi Koefisien Regresi Secara Individu
 Uji ini digunakan untuk menguji ada tidaknya pengaruh masing-masing variabel prediktor terhadap variabel respon. Langkah-langkahnya:
 - a. Hipotesis
 $H_0: \beta_j = 0$, (koefisien regresi X_j tidak signifikan)
 $H_1: \beta_j \neq 0, j = 1, 2, \dots, k$ (koefisien regresi X_j signifikan)
 - b. Statistik uji

$$t_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \text{ dengan } se(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}}$$
 dimana C_{jj} adalah elemen diagonal dari $(\mathbf{X}'\mathbf{X})^{-1}$ dan $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-k-1}$
 - c. Kriteria uji
 H_0 ditolak jika $|t_0| > t_{\text{tabel}} = t_{(\alpha/2; n-k-1)}$ atau $p\text{-value} < \alpha$.

2.1.3 Ukuran Kecocokan Model Regresi

Kecocokan model regresi dapat diukur dengan melihat nilai-nilai berikut:

1. Koefisien Determinasi yang Disesuaikan (*Adjusted R²*)
 Pada model regresi berganda menurut Gujarati (2004), R^2 dikenal untuk mengukur proporsi atau persentase total variasi dalam Y yang dijelaskan oleh model regresi.

$$R_{\text{adj}}^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / (n-k-1)}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / (n-1)}$$

2. *Residual Standard Error (RSE)*
Residual Standard Error (RSE) adalah estimasi simpangan baku (*standard deviation*) dari ε . ε merupakan selisih dari nilai yang diprediksi dengan nilai yang diamati. Secara umum, menurut James *et al.* (2013), RSE didefinisikan sebagai berikut:

$$RSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-k-1}} = \sqrt{MSE}$$

2.2 Multikolinieritas

Tidak adanya multikolinieritas atau hubungan antar variabel prediktor merupakan salah satu asumsi yang harus dipenuhi dalam model regresi berganda. Menurut Widarjono (2005), dampak adanya multikolinieritas, jika menggunakan teknik estimasi dengan metode OLS tetapi masih mempertahankan asumsi lain adalah sebagai berikut:

1. Estimator masih bersifat BLUE dengan adanya multikolinieritas, namun estimator mempunyai varian yang besar sehingga sulit mendapatkan estimasi yang tepat.
2. Akibat nomor 1, maka interval estimasi akan cenderung lebih lebar dan nilai hitung statistik uji t akan kecil, sehingga membuat variabel prediktor secara statistik tidak signifikan mempengaruhi variabel respon.
3. Walaupun secara individu variabel prediktor tidak berpengaruh terhadap variabel respon melalui uji statistik t , namun nilai koefisien determinasi (R^2) masih relatif tinggi.

Ada beberapa cara yang dapat digunakan untuk mendeteksi masalah multikolinieritas dalam suatu model regresi, yaitu:

1. Jika nilai R^2 tinggi (misalnya: antara 0,7 dan 1), tetapi hanya sedikit variabel prediktor yang signifikan (Gujarati, 2004).
2. Jika koefisien korelasi mendekati satu atau lebih besar dari 0,75 maka dapat diduga ada multikolinieritas dalam model (Marcus *et al.*, 2012). Koefisien korelasi parsial

antar variabel prediktor dapat dihitung dengan rumus korelasi *pearson* sebagai berikut:

$$r_{X_j X_l} = \frac{n \sum X_j X_l - \sum X_j \sum X_l}{\sqrt{n \sum X_j^2 - (\sum X_j)^2} \sqrt{n \sum X_l^2 - (\sum X_l)^2}} \text{ dengan } j = 1, 2, \dots, k \text{ dan } l = 1, 2, \dots, k.$$

3. Jika nilai VIF lebih besar dari 5 atau 10 telah mengindikasikan adanya multikolinieritas pada model regresi (Montgomery *et al.*, 2012). *Variance Inflation Factors* (VIF) dapat dihitung dengan menggunakan rumus sebagai berikut:

$$VIF = \frac{1}{1 - R_j^2} \text{ dengan } R_j^2 \text{ adalah koefisien determinasi ke-} j \text{ dimana } j = 1, 2, \dots, k.$$

Menurut Widarjono (2005), ada beberapa cara yang bisa digunakan untuk mengatasi kasus multikolinieritas antara lain:

1. Menghilangkan variabel prediktor
2. Menambahkan data
3. Mentransformasi variabel

Selain itu, multikolinieritas dapat juga ditangani dengan menggunakan analisis regresi *ridge* dan regresi komponen utama (Montgomery *et al.*, 2012).

2.3 Deteksi *Outlier*

Menurut Anisa (2010), terdapat beberapa kemungkinan adanya data *outlier*, yaitu *outlier* pada variabel prediktor, *outlier* pada variabel respon, atau keduanya. Adanya *outlier* pada variabel prediktor dapat dideteksi dengan menghitung jarak mahalanobis. Untuk mengukur jarak mahalanobis digunakan vektor rata-rata dan matriks kovarian. Sebuah pengamatan X_i dideteksi sebagai *outlier* jika jarak mahalanobisnya:

$$d_{MD}^2 = (\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) > \chi_{(k, \alpha)}^2$$

dengan $\bar{\mathbf{X}}$ dan \mathbf{S} adalah vektor rata-rata dan matriks kovarian dari data.

Selanjutnya, untuk mendeteksi adanya *outlier* pada variabel respon dapat dilakukan dengan melihat residual (*error*) dari model regresi (Anisa, 2010). Salah satu metode yang dapat digunakan adalah metode *DFFITs* (*Difference in Fit Standardized*). Perhitungan $DFFITs_i$ adalah sebagai berikut:

$$DFFITs_i = t_i \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{\frac{1}{2}} ; \quad t_i = e_i \sqrt{\frac{n - p - 1}{SSE(1 - h_{ii}) - e_i^2}}$$

dimana

e_i : residual ke- i

h_{ii} : elemen baris ke- i kolom ke- i dari matriks $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

Data dikatakan *outlier* apabila nilai $|DFFITs_i| > 2 \sqrt{\frac{p}{n}}$ dengan p adalah banyaknya parameter dan n adalah banyaknya pengamatan (Perihatini, 2018).

2.4 Regresi Komponen Utama

Menurut Soemartini (2008), ada beberapa keuntungan dari regresi komponen utama apabila dibandingkan metode-metode lain dalam mengatasi masalah multikolinieritas, diantaranya:

1. Dapat menghilangkan korelasi secara bersih (korelasi = 0) sehingga masalah multikolinieritas dapat benar-benar teratasi secara bersih.
2. Dapat digunakan untuk segala kondisi data/penelitian.
3. Dapat digunakan tanpa mengurangi jumlah variabel asal.

4. Meskipun regresi komponen utama ini memiliki tingkat kesulitan yang tinggi, akan tetapi kesimpulan yang diberikan lebih akurat dibandingkan dengan metode lain.

Menurut Notiragayu dan Nisa (2008), ada dua tahapan dalam regresi komponen utama, yaitu tahap pertama adalah melakukan analisis komponen utama dan tahap kedua adalah meregresikan komponen-komponen utama dengan variabel respon.

2.4.1 Analisis Komponen Utama

Pada analisis komponen utama, didasarkan bahwa skala pengukuran dari X_1, X_2, \dots, X_k sama, kemudian dibentuk variabel baru W yang disebut sebagai komponen utama yang merupakan kombinasi linier dari X_1, X_2, \dots, X_k dengan bentuk sebagai berikut (Johnson dan Wichern, 2007):

$$\begin{aligned} W_1 &= \mathbf{a}_1' \mathbf{X} = a_{11} X_1 + a_{12} X_2 + \dots + a_{1k} X_k \\ W_2 &= \mathbf{a}_2' \mathbf{X} = a_{21} X_1 + a_{22} X_2 + \dots + a_{2k} X_k \\ &\vdots \\ W_k &= \mathbf{a}_k' \mathbf{X} = a_{k1} X_1 + a_{k2} X_2 + \dots + a_{kk} X_k \end{aligned}$$

dengan

$$\begin{aligned} \text{Cov}(W_j, W_l) &= \mathbf{a}_j' \mathbf{S} \mathbf{a}_l & j, l = 1, 2, \dots, k \\ \text{Var}(W_j) &= \mathbf{a}_j' \mathbf{S} \mathbf{a}_j & j = 1, 2, \dots, k \end{aligned}$$

Syarat untuk membentuk komponen utama yang merupakan kombinasi linier dari X_1, X_2, \dots, X_k agar mempunyai varian maksimum adalah dengan memilih *eigenvector* yaitu $\mathbf{a}_j' = [a_{1j}, a_{2j}, \dots, a_{kj}]$ dengan $j = 1, 2, \dots, k$, sedemikian sehingga $\text{Var}(W_j) = \mathbf{a}_j' \mathbf{S} \mathbf{a}_j$ maksimum dengan kendala $\mathbf{a}_j' \mathbf{a}_j = \mathbf{1}$.

Untuk mencari *eigenvector* yaitu $\mathbf{a}_j' = [a_{1j}, a_{2j}, \dots, a_{kj}]$ dari \mathbf{S} berlaku $(\mathbf{S} - \lambda \mathbf{I}) \mathbf{a}_j = \mathbf{0}$. Akan tetapi, sebelumnya dicari terlebih dahulu *eigenvalue* (λ) dari \mathbf{S} dengan syarat $|\mathbf{S} - \lambda \mathbf{I}| = 0$.

Keragaman yang dapat dijelaskan oleh komponen utama ke- j terhadap keragaman total adalah

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_k} \times 100\% ; j = 1, 2, \dots, k$$

Sedangkan secara kumulatif, keragaman total yang dijelaskan oleh m komponen utama adalah

$$\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^k \lambda_j} \times 100\% \text{ dengan } m \leq k$$

m adalah banyaknya komponen utama yang terpilih.

Pada masalah regresi biasanya skala pengukuran untuk variabel-variabel prediktor X_1, X_2, \dots, X_k biasanya belum sama, sehingga perlu disamakan dengan cara mentransformasi ke dalam variabel baku Z (Widiharih, 2001).

Variabel baku Z diperoleh dari transformasi terhadap variabel asal yang dirumuskan sebagai berikut:

$$\begin{aligned} Z_1 &= \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}} \\ Z_2 &= \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}} \\ &\vdots \\ Z_k &= \frac{(X_k - \mu_k)}{\sqrt{\sigma_{kk}}} \end{aligned}$$

Menurut Johnson dan Wichern (2007), komponen utama yang dibentuk sebagai kombinasi linier dari variabel yang dibakukan yaitu Z_1, Z_2, \dots, Z_k adalah:

$$W_1 = \mathbf{a}_1' \mathbf{Z} = a_{11} Z_1 + a_{12} Z_2 + \dots + a_{1k} Z_k$$

$$W_2 = \mathbf{a}_2' \mathbf{Z} = a_{21} Z_1 + a_{22} Z_2 + \dots + a_{2k} Z_k$$

⋮

$$W_k = \mathbf{a}_k' \mathbf{Z} = a_{k1} Z_1 + a_{k2} Z_2 + \dots + a_{kk} Z_k$$

dengan Cov (\mathbf{Z}) adalah matriks korelasi \mathbf{R} yang didefinisikan sebagai berikut:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1k} \\ r_{21} & r_{22} & \dots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & r_{kk} \end{bmatrix}$$

Nilai r_{ij} dapat dihitung dengan rumus korelasi *pearson* berikut:

$$r_{ij} = r_{Z_i Z_j} = \frac{n \sum Z_i Z_j - \sum Z_i \sum Z_j}{\sqrt{n \sum Z_i^2 - (\sum Z_i)^2} \sqrt{n \sum Z_j^2 - (\sum Z_j)^2}}$$

Semua formula yang telah dijelaskan berdasarkan variabel-variabel X_1, X_2, \dots, X_k dengan matriks \mathbf{S} akan berlaku untuk variabel-variabel Z_1, Z_2, \dots, Z_k dengan matriks \mathbf{R} .

Syarat untuk membentuk komponen utama yang merupakan kombinasi linier dari Z_1, Z_2, \dots, Z_k agar mempunyai varian maksimum adalah dengan memilih *eigenvector* yaitu $\mathbf{a}'_j = [\mathbf{a}_{1j}, \mathbf{a}_{2j}, \dots, \mathbf{a}_{kj}]$ dengan $j = 1, 2, \dots, k$, sedemikian sehingga $\text{Var}(W_j) = \mathbf{a}'_j \mathbf{R} \mathbf{a}_j$ maksimum dengan kendala $\mathbf{a}'_j \mathbf{a}_j = 1$.

Untuk mencari *eigenvector* yaitu $\mathbf{a}'_j = [\mathbf{a}_{1j}, \mathbf{a}_{2j}, \dots, \mathbf{a}_{kj}]$ dari \mathbf{R} berlaku $(\mathbf{R} - \lambda \mathbf{I}) \mathbf{a}_j = \mathbf{0}$. Akan tetapi, sebelumnya dicari terlebih dahulu *eigenvalue* (λ) dari \mathbf{R} dengan syarat $|\mathbf{R} - \lambda \mathbf{I}| = 0$.

Menurut Johnson dan Wichern (2007), keragaman total yang dapat dijelaskan oleh komponen utama ke- j dari variabel yang dibakukan yaitu Z_1, Z_2, \dots, Z_k adalah

$$\frac{\lambda_j}{k} \times 100\% ; j = 1, 2, \dots, k$$

Pemilihan komponen utama yang digunakan untuk regresi komponen utama dapat dilakukan dengan cara-cara berikut:

1. Menggunakan kumulatif proporsi keragaman total yang mampu dijelaskan oleh komponen-komponen utama yang dipilih. Marison (1978) dalam Sriningsih *et al.* (2018) menyarankan untuk memilih komponen-komponen utama yang mempunyai kumulatif proporsi keragaman total 75%.
2. Menggunakan *eigenvalue* yang mempunyai nilai lebih besar atau sama dengan satu (Safidah, 2014).
3. Melihat *scree plot* yaitu plot antara *eigenvalue* λ_j dengan j . Penentuan banyaknya komponen utama yang digunakan adalah dengan memperhatikan patahan siku dari *scree plot* (Jumiati, 2018).

Meskipun demikian, sebagian ahli menyarankan agar memilih komponen utama yang mempunyai *eigenvalue* lebih besar dari satu. Hal ini dikarenakan, jika *eigenvalue* kurang dari satu, maka keragaman data yang diterangkan semakin kecil (Sringingsih *et al.*, 2018).

Setelah komponen utama diperoleh, langkah berikutnya adalah menghitung skor komponen utama dari setiap pengamatan dengan rumus (Widiharih, 2001):

$$SK-W_{ji} = \mathbf{a}'_j \mathbf{Z}_i \text{ untuk } i = 1, 2, \dots, n \text{ dan } j = 1, 2, \dots, m$$

dimana $SK-W_{ji}$: skor komponen utama ke-j untuk pengamatan ke-i

\mathbf{a}'_j : *eigenvector* komponen utama ke-j

\mathbf{Z}_i : vektor skor variabel baku yang diamati pada pengamatan ke-i

2.4.2 Analisis Regresi Komponen Utama

Setelah dilakukan analisis komponen utama dan telah diperoleh skor komponen utama, langkah selanjutnya adalah meregresikan komponen-komponen utama terpilih (\mathbf{W}) dengan variabel respon menggunakan metode *Ordinary Least Squares* (OLS). Bentuk persamaan regresi komponen utama sama seperti bentuk persamaan regresi pada umumnya, yaitu:

$$Y_i = \alpha_0 + \alpha_1 W_{1i} + \alpha_2 W_{2i} + \dots + \alpha_m W_{mi} + \gamma_i$$

Atau dapat juga dituliskan dalam bentuk matriks sebagai berikut:

$$\mathbf{y} = \mathbf{W}\boldsymbol{\alpha} + \boldsymbol{\gamma}$$

dengan

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; \mathbf{W} = \begin{bmatrix} 1 & W_{11} & W_{21} & \dots & W_{m1} \\ 1 & W_{12} & W_{22} & \dots & W_{m2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_{1n} & W_{2n} & \dots & W_{mn} \end{bmatrix}; \boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}; \boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix}$$

Parameter $\hat{\boldsymbol{\alpha}}$ dapat diestimasi menggunakan metode *Ordinary Least Squares* (OLS) menggunakan rumus:

$$\hat{\boldsymbol{\alpha}} = (\mathbf{W}'\mathbf{W})^{-1}(\mathbf{W}'\mathbf{y})$$

2.5 Metode Minimum Covariance Determinant (MCD)

Metode *Minimum Covariance Determinant* (MCD) merupakan metode untuk mendapatkan penaksir yang *robust* yang didapatkan dari rata-rata dan kovarian dari sebagian pengamatan yang memiliki determinan matriks kovarian yang minimum. Metode ini bertujuan untuk menemukan *h-observations* atau subsampel dari data, dimana matriks kovariannya memiliki determinan terkecil.

Misalkan n adalah banyaknya pengamatan dan k adalah banyaknya variabel, maka algoritma FAST-MCD secara garis besar menurut Rousseuw dan Driessen (1999) adalah sebagai berikut:

1. Menentukan $h = \left\lfloor \frac{n+k+1}{2} \right\rfloor$, tetapi dapat juga memilih h dengan $\left\lfloor \frac{n+k+1}{2} \right\rfloor \leq h \leq n$.
2. Jika $h = n$, maka perkiraan lokasi $\bar{\mathbf{X}}_{MCD}$ adalah rata-rata dari himpunan data tersebut dan perkiraan \mathbf{S}_{MCD} adalah matriks kovariannya.
3. Untuk $h < n$ dan $k \geq 2$. Jika n kecil (misalkan $n \leq 600$), maka
 - a. Bentuk h himpunan bagian H_1 awal dengan cara:
 - 1) Ambil secara acak $(k+1)$ himpunan bagian J dan hitung $\bar{\mathbf{X}}_0 =$ rata-rata (J) dan $\mathbf{S}_0 = \text{cov}(J)$ [Jika $\det(\mathbf{S}_0) = 0$, maka perluas J dengan menambah satu demi satu pengamatan acak lainnya sampai $\det(\mathbf{S}_0) > 0$].
 - 2) Hitung jarak $d_0^2(i) = (\mathbf{X}_i - \bar{\mathbf{X}}_0)' \mathbf{S}_0^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}_0)$ untuk $i = 1, 2, \dots, n$. Urutkan ke dalam $d_0(\pi(1)) \leq \dots \leq d_0(\pi(n))$ dan tempatkan pada $H_1 = \{ \pi(1), \dots, \pi(h) \}$

- b. Lakukan *C-Steps* yaitu dengan cara:
 - 1) Hitung $\bar{\mathbf{X}}_1 = \frac{1}{h} \sum_{i \in H_1} \mathbf{X}_i$
 - 2) Hitung $\mathbf{S}_1 = \frac{1}{h} \sum_{i \in H_1} (\mathbf{X}_i - \bar{\mathbf{X}}_1)(\mathbf{X}_i - \bar{\mathbf{X}}_1)'$
 - 3) Jika $\det(\mathbf{S}_1) \neq 0$, maka hitung jarak relatif dengan rumus jarak mahalanobis, sebagai berikut:

$$d_1^2(i) = (\mathbf{X}_i - \bar{\mathbf{X}}_1)' \mathbf{S}_1^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}_1) \text{ untuk } i = 1, 2, \dots, n$$
 - 4) Urutkan ke dalam $d_1(\pi(1)) \leq \dots \leq d_1(\pi(n))$
 - 5) Tempatkan pada $H_2 = \{ \pi(1), \dots, \pi(h) \}$
 - 6) Hitung $\bar{\mathbf{X}}_2 = \text{rata-rata}(H_2)$ dan $\mathbf{S}_2 = \text{cov}(H_2)$
- c. Ulangi *C-Steps* sehingga diperoleh $\det(\mathbf{S}_1) \geq \det(\mathbf{S}_2) \geq \dots \geq \det(\mathbf{S}_{m-1}) \geq \det(\mathbf{S}_m)$ yang bernilai nonnegatif.
- d. Perulangan berhenti jika $\det(\mathbf{S}_m) = 0$ atau $\det(\mathbf{S}_m)$ konvergen dengan $\det(\mathbf{S}_{m-1})$.
- e. Laporkan solusi $(\bar{\mathbf{X}}, \mathbf{S})$ dengan $\det(\mathbf{S})$ terkecil.

2.6 Metode Least Trimmed Squares (LTS)

Metode *Least Trimmed Squares* (LTS) merupakan suatu metode pendugaan parameter regresi *robust* untuk meminimumkan jumlah kuadrat h residual (fungsi objektif) (Rousseeuw dan Leroy, 1987). Fungsi objektif dari metode LTS adalah sebagai berikut:

$$Q = \sum_{i=1}^h e_{(i:n)}^2$$

dengan

$$h = \left\lfloor \frac{n+p+1}{2} \right\rfloor \text{ atau } h = \gamma * n \text{ dimana } \gamma = (1 - \alpha)$$

γ : persentase besar data yang akan dipangkas (*trimmed*)

α : nilai breakdown (persentase banyaknya pencilan terhadap seluruh data)

$$e_i = (Y_i - \hat{Y}_i)$$

$$e_1^2 \leq e_2^2 \leq \dots \leq e_n^2 : \text{kuadrat residual diurutkan dari terkecil ke terbesar}$$

: banyaknya pengamatan

: banyaknya parameter

Jumlah h menunjukkan sejumlah subset data dengan kuadrat residual terkecil. Menurut Rousseeuw dan Driessen (1999) dalam Willems dan Aels (2005), estimasi parameter LTS dilakukan dengan menggabungkan algoritma FAST-LTS dan *C-Steps* seperti berikut:

1. Hitung estimasi awal parameter regresi yaitu $\hat{\boldsymbol{\beta}}^0$.
2. Hitung residual $e_0(i)$ untuk $i = 1, 2, \dots, n$.
3. Hitung residual $e_0^2(i)$ untuk $i = 1, 2, \dots, n$.
4. Urutkan ke dalam $e_0^2(\pi(1)) \leq \dots \leq e_0^2(\pi(n))$ dan tempatkan pada $H_1 = \{ \pi(1), \dots, \pi(h) \}$ dengan $h = \gamma * n$
5. Lakukan *C-Steps* yaitu dengan cara:
 - a. Hitung $\hat{\boldsymbol{\beta}}^1$, yaitu estimasi parameter regresi dari H_1 .
 - b. Hitung residual $e_1(i)$ untuk $i = 1, 2, \dots, n$.
 - c. Hitung residual $e_1^2(i)$ untuk $i = 1, 2, \dots, n$.
 - d. Hitung $Q_1 = \sum_{i \in H_1} e_1^2(i)$.
 - e. Urutkan ke dalam $e_1^2(\pi(1)) \leq \dots \leq e_1^2(\pi(n))$ dan tempatkan pada $H_2 = \{ \pi(1), \dots, \pi(h) \}$ dengan $h = \gamma * n$.
 - f. Hitung $\hat{\boldsymbol{\beta}}^2$, yaitu estimasi parameter regresi dari H_2 .

6. Ulangi *C-Steps* sehingga diperoleh $Q_1 \geq Q_2 \geq \dots \geq Q_{m-1} \geq Q_m$ yang bernilai nonnegatif.
7. Perulangan berhenti jika $Q_m \leq Q_{m-1}$ dan nilai estimasi parameter regresi $\hat{\beta}^m$ konvergen dengan $\hat{\beta}^{m-1}$.
8. Laporkan solusi $\hat{\beta}$ dengan fungsi objektif Q terkecil.

2.7 Indeks Pembangunan Manusia (IPM)

Indeks Pembangunan Manusia (IPM) merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia. IPM dibangun dari tiga dimensi dasar yang digunakan sebagai ukuran kualitas hidup manusia. Dimensi tersebut mencakup kesehatan, pendidikan, dan ekonomi.

Adapun variabel-variabel yang dapat digunakan untuk mengukur masing-masing dimensi dasar IPM tersebut adalah sebagai berikut:

1. Untuk mengukur dimensi kesehatan dapat digunakan banyaknya sarana kesehatan (Putra dan Ratnasari, 2015).
2. Untuk mengukur dimensi pendidikan dapat digunakan angka partisipasi kasar dan angka partisipasi murni (Ekosiswoyo *et al.*, 2008).
3. Untuk mengukur dimensi ekonomi dapat digunakan presentase penduduk miskin, jumlah penduduk usia 15 tahun ke atas yang bekerja, dan tingkat partisipasi angkatan kerja (Putra dan Ratnasari, 2015).

3. METODE PENELITIAN

3.1 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data Indeks Pembangunan Manusia (IPM), tingkat partisipasi angkatan kerja, angka partisipasi sekolah, presentase penduduk miskin, jumlah penduduk usia 15 tahun ke atas yang bekerja, banyaknya sarana kesehatan, angka partisipasi kasar, dan angka partisipasi murni. Data-data tersebut merupakan data sekunder yang diperoleh dari publikasi Badan Pusat Statistik Provinsi Jawa Tengah dan Dinas Kesehatan Provinsi Jawa Tengah tahun 2017.

3.2 Variabel Penelitian

- Y : Indeks Pembangunan Manusia (IPM)
- X_1 : Tingkat Partisipasi Angkatan Kerja (TPAK)
- X_2 : Angka partisipasi sekolah (APS)
- X_3 : Persentase penduduk miskin
- X_4 : Jumlah penduduk usia 15 tahun ke atas yang bekerja
- X_5 : Banyaknya sarana kesehatan
- X_6 : Angka Partisipasi Kasar (APK)
- X_7 : Angka Partisipasi Murni (APM)

3.3 Tahapan Analisis Data

1. Mencari data yang akan digunakan, meliputi data IPM, tingkat partisipasi angkatan kerja, angka partisipasi sekolah, presentase penduduk miskin, jumlah penduduk usia 15 tahun ke atas yang bekerja, banyaknya sarana kesehatan, angka partisipasi kasar, dan angka partisipasi murni.
2. Menentukan variabel respon dan variabel prediktor dari data yang diperoleh.
3. Mempersiapkan *software R 3.4.3*.
4. Menginstal beberapa *package* dalam *software R 3.4.3* diantaranya yaitu *lmtest*, *car*, *MASS*, dan *robustbase*.

5. Memasukkan data ke dalam *software R 3.4.3*.
6. Melakukan analisis deskriptif terhadap data penelitian.
7. Melakukan standarisasi data penelitian.
8. Melakukan analisis regresi berganda dengan metode OLS.
9. Melakukan pemeriksaan asumsi non-multikolinieritas dengan cara melihat nilai VIF dan nilai koefisien korelasi *pearson* antar variabel prediktor.
10. Mendeteksi adanya *outlier* pada variabel prediktor menggunakan jarak mahalanobis.
11. Melakukan analisis komponen utama klasik dan analisis komponen utama *robust* terhadap variabel prediktor. Langkah-langkah analisis komponen utama klasik dan analisis komponen utama *robust* adalah sebagai berikut:
 - A. Analisis komponen utama klasik, dilakukan dengan cara berikut:
 - a) Menghitung penaksir klasik, yaitu matriks korelasi antar variabel prediktor yang sudah distandarisasi.
 - b) Mencari *eigenvalue* dan *eigenvector*.
 - c) Menentukan komponen-komponen utama yang akan digunakan, yaitu komponen utama yang mempunyai *eigenvalue* lebih besar atau sama dengan satu.
 - d) Menghitung proporsi kumulatif varian yang dapat dijelaskan oleh komponen-komponen utama yang terpilih.
 - e) Menghitung skor komponen utama.
 - B. Analisis komponen utama *robust*, dilakukan dengan cara berikut:
 - a) Menghitung penaksir *robust*, yaitu matriks kovarian antar variabel prediktor yang sudah distandarisasi dengan metode C-Step melalui algoritma FAST-MCD.
 - b) Mencari *eigenvalue* dan *eigenvector*.
 - c) Menentukan komponen-komponen utama yang akan digunakan, yaitu komponen utama yang mempunyai *eigenvalue* lebih besar atau sama dengan satu.
 - d) Menghitung proporsi kumulatif varian yang dapat dijelaskan oleh komponen-komponen utama yang terpilih.
 - e) Menghitung skor komponen utama.
12. Meregresikan komponen-komponen utama terpilih dari analisis komponen utama klasik dan analisis komponen utama *robust* dengan variabel respon menggunakan metode OLS, sehingga diperoleh dua buah model yaitu model RKU Klasik-OLS dan model RKU *Robust*-OLS.
13. Melakukan uji signifikansi model regresi melalui uji F.
14. Melakukan uji koefisien regresi secara individu melalui uji t.
15. Melakukan deteksi multikolinieritas dan diperoleh hasil bahwa masalah multikolinieritas sudah teratasi.
16. Menghitung nilai *Adjusted R²*.
17. Menghitung nilai RSE.
18. Mendeteksi adanya *outlier* pada dua buah model yang diperoleh dari langkah 12 menggunakan metode *DFFITs*.
19. Jika ada *outlier*, maka lakukan analisis regresi komponen utama *robust*, yaitu meregresikan komponen-komponen utama terpilih dengan variabel respon menggunakan metode LTS. Sehingga, diperoleh dua buah model lagi, yaitu RKU Klasik-LTS dan RKU *Robust*-LTS.
20. Melakukan uji signifikansi model regresi melalui uji F.

21. Melakukan uji koefisien regresi secara individu melalui uji t.
22. Menghitung nilai *Adjusted R²*.
23. Menghitung nilai RSE.
24. Mengembalikan persamaan regresi ke bentuk variabel standar.
25. Mengembalikan persamaan regresi ke bentuk variabel asal.
26. Menarik kesimpulan.

4. HASIL DAN PEMBAHASAN

4.1 Deskripsi Data Penelitian

Disajikan deskripsi data penelitian dalam Tabel 1 sebagai berikut:

Tabel 1. Deskripsi Data Penelitian

Variabel	Minimum	Maksimum	Rata-Rata	Standar Deviasi
Y	64,86	82,01	71,19	4,48
X ₁	64,48	76,37	69,17	2,95
X ₂	53,72	90,74	70,20	9,81
X ₃	4,62	20,32	12,49	4,13
X ₄	58582	899796	491047,83	201784,47
X ₅	11	61	31,40	11,66
X ₆	52,98	109,61	85,58	14,54
X ₇	42,18	75,72	59,50	7,91

4.2 Analisis Regresi Berganda

Pada awal penelitian dilakukan analisis regresi berganda dengan estimasi parameter menggunakan *Ordinary Least Squares* (OLS). Karena terdapat perbedaan unit satuan pada variabel-variabel tersebut, maka dalam penelitian ini dilakukan standarisasi terhadap variabel respon dan variabel prediktor terlebih dahulu.

4.2.1 Estimasi Parameter Model

Dari hasil *output* didapatkan model regresi linier berganda sebagai berikut:

$$\hat{Q} = 0,08068 Z_1 + 0,56930 Z_2 - 0,45329 Z_3 - 0,19787 Z_4 + 0,28958 Z_5 + 0,42591 Z_6 - 0,47257 Z_7$$

Keterangan:

\hat{Q} : Data Indeks Pembangunan Manusia (IPM) yang Distandarisasi

Z₁ : Data Tingkat Partisipasi Angkatan Kerja (TPAK) yang Distandarisasi

Z₂ : Data Angka Partisipasi Sekolah (APS) yang Distandarisasi

Z₃ : Data Presentase Penduduk Miskin yang Distandarisasi

Z₄ : Data Jumlah Penduduk Usia 15 Tahun ke Atas yang Bekerja yang Distandarisasi

Z₅ : Data Banyaknya Sarana Kesehatan yang Distandarisasi

Z₆ : Data Angka Partisipasi Kasar (APK) yang Distandarisasi

Z₇ : Data Angka Partisipasi Murni (APM) yang Distandarisasi

4.2.2 Uji Signifikansi dalam Regresi Berganda

Setelah dilakukan estimasi parameter model, untuk mengukur ketepatan model maka dilakukan uji signifikansi model regresi dan uji signifikansi koefisien regresi secara individu seperti berikut:

1. Uji Signifikansi Model Regresi
 - a. Hipotesis
 $H_0: \beta_1 = \beta_2 = \dots = \beta_7 = 0$ (Model regresi tidak sesuai)
 $H_1: \text{ada minimal satu } \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, 7$ (Model regresi sesuai)
 - b. Statistik uji
 $F_0 = 15,21$ dan $p\text{-value} = 4,965 \times 10^{-8}$
 - c. Kriteria uji
 H_0 ditolak jika $F_0 > F_{(0,05;7;28)} = 2,359$ atau $p\text{-value} < \alpha = 0,05$.
 - d. Kesimpulan
 H_0 ditolak sehingga dapat disimpulkan bahwa pada taraf signifikansi 5%, model sesuai yang berarti ada hubungan linier antara variabel respon dengan variabel predictor atau model sesuai.
2. Uji Signifikansi Koefisien Regresi Secara Individu
 - a. Hipotesis
 $H_0: \beta_j = 0$, (koefisien regresi X_j tidak signifikan)
 $H_1: \beta_j \neq 0, j = 1, 2, \dots, 7$ (koefisien regresi X_j signifikan)
 - b. Statistik uji
 t_0 dan $p\text{-value}$ yang disajikan dalam Tabel 2.
 - c. Kriteria uji
 H_0 ditolak jika $|t_0| > t_{(0,025;28)} = 2,048$ atau $p\text{-value} < \alpha = 0,05$.
 - d. Kesimpulan

Tabel 2. Uji t Regresi Berganda

Variabel	$ t_0 $	$p\text{-value}$	Keputusan	Kesimpulan
Z_1	0,780	0,441875	H_0 diterima	Tidak Signifikan
Z_2	2,735	0,010692	H_0 ditolak	Signifikan
Z_3	4,106	0,000316	H_0 ditolak	Signifikan
Z_4	1,181	0,247586	H_0 diterima	Tidak Signifikan
Z_5	1,750	0,091064	H_0 diterima	Tidak Signifikan
Z_6	2,268	0,031262	H_0 ditolak	Signifikan
Z_7	2,318	0,027992	H_0 ditolak	Signifikan

Dari Tabel tersebut, pada taraf signifikansi 5%, variabel angka partisipasi sekolah (Z_2), presentase penduduk miskin (Z_3), angka partisipasi kasar (Z_6), dan angka partisipasi murni (Z_7) berpengaruh signifikan terhadap IPM (Q). Sedangkan variabel tingkat partisipasi angkatan kerja (Z_1), jumlah penduduk usia 15 tahun ke atas yang bekerja (Z_4), dan banyaknya sarana kesehatan (Z_5) tidak berpengaruh signifikan terhadap IPM (Q).

4.2.3 Ukuran Kecocokan Model Rgresi

Diperoleh nilai RSE sebesar 0,5028 dan *Adjusted* R^2 sebesar 0,7397 yang menunjukkan bahwa besar pengaruh tingkat partisipasi angkatan kerja, angka partisipasi sekolah, presentase penduduk miskin, jumlah penduduk usia 15 tahun ke atas yang bekerja, banyaknya sarana kesehatan, angka partisipasi kasar, dan angka partisipasi murni terhadap

IPM adalah 73,97%. Sedangkan sisanya sebesar 26,03% IPM dipengaruhi oleh faktor lain di luar model.

4.3 Deteksi Multikolinieritas

1. Menggunakan nilai *Variance Inflation Factors* (VIF)

Jika nilai VIF lebih besar dari 5 atau 10 dapat mengindikasikan adanya multikolinieritas dalam model regresi.

Tabel 3. Deteksi Multikolinieritas dengan VIF

Variabel	VIF	Keterangan
Z ₁	1,438382	VIF < 5
Z ₂	5,824945	VIF > 5
Z ₃	1,638559	VIF < 5
Z ₄	3,775565	VIF < 5
Z ₅	3,681816	VIF < 5
Z ₆	4,743491	VIF < 5
Z ₇	5,590096	VIF > 5

Dilihat dari Tabel 3, variabel angka partisipasi sekolah (Z₂) dan angka partisipasi murni (Z₇) mempunyai nilai VIF lebih besar dari 5 yang mengindikasikan adanya multikolinieritas dalam model regresi.

2. Menggunakan nilai koefisien korelasi parsial antar variabel prediktor

Jika nilai koefisien korelasi parsial antar variabel prediktor mendekati satu atau lebih besar dari 0,75 maka dapat diduga ada multikolinieritas dalam model regresi. Pada *output* diperoleh matriks korelasi (**R**) antar variabel prediktor berikut:

$$\mathbf{R} = \begin{bmatrix} 1,000 & -0,223 & -0,073 & 0,073 & -0,132 & -0,401 & -0,304 \\ -0,223 & 1,000 & -0,332 & -0,345 & -0,122 & 0,790 & 0,823 \\ -0,073 & -0,332 & 1,000 & 0,261 & 0,298 & -0,447 & -0,234 \\ 0,073 & -0,345 & 0,261 & 1,000 & 0,803 & -0,160 & 0,009 \\ -0,132 & -0,122 & 0,298 & 0,803 & 1,000 & 0,053 & 0,223 \\ -0,401 & 0,790 & -0,447 & -0,160 & 0,053 & 1,000 & 0,801 \\ -0,304 & 0,823 & -0,234 & 0,009 & 0,223 & 0,801 & 1,000 \end{bmatrix}$$

Dari matriks tersebut, terlihat bahwa ada korelasi yang cukup tinggi antara variabel angka partisipasi sekolah (Z₂) dengan angka partisipasi kasar (Z₆) sebesar 0,790. Angka partisipasi sekolah (Z₂) dengan angka partisipasi murni (Z₇) sebesar 0,823. Jumlah penduduk usia 15 tahun ke atas yang bekerja (Z₄) dengan banyaknya sarana kesehatan (Z₅) sebesar 0,803. Angka partisipasi kasar (Z₆) dengan angka partisipasi murni (Z₇) sebesar 0,801. Karena nilai-nilai koefisien korelasi parsial tersebut lebih besar dari 0,75, maka hal ini mengindikasikan adanya multikolinieritas dalam model regresi.

Dari pendeteksian multikolinieritas dapat disimpulkan bahwa ada gejala multikolinieritas pada model regresi sehingga pada penelitian ini akan dilakukan tindakan lanjutan untuk mengatasi multikolinieritas dengan menggunakan regresi komponen utama. Regresi komponen utama merupakan metode yang menggabungkan analisis komponen utama dengan analisis regresi.

4.4 Analisis Komponen Utama

Sebelum melakukan analisis komponen utama, terlebih dahulu dilakukan pendeteksian *outlier* pada variabel prediktor menggunakan jarak mahalanobis. Dengan nilai $\chi^2_{(7;0,05)} = 14,067$ dari *output* terlihat bahwa ada dua *outlier* yang terdeteksi, yaitu data ke-32 dan 33 dengan nilai d^2_{MD} masing-masing adalah 16,736 dan 17,507. Karena terdapat *outlier* pada variabel prediktor, maka pada penelitian ini, analisis komponen utama dilakukan dengan dua cara, yaitu analisis komponen utama klasik dan analisis komponen utama *robust*.

4.4.1 Analisis Komponen Utama Klasik

Dengan menggunakan penaksir klasik, yaitu matriks korelasi (**R**) antar variabel prediktor berikut:

$$\mathbf{R} = \begin{bmatrix} 1,000 & -0,223 & -0,073 & 0,073 & -0,132 & -0,401 & -0,304 \\ -0,223 & 1,000 & -0,332 & -0,345 & -0,122 & 0,790 & 0,823 \\ -0,073 & -0,332 & 1,000 & 0,261 & 0,298 & -0,447 & -0,234 \\ 0,073 & -0,345 & 0,261 & 1,000 & 0,803 & -0,160 & 0,009 \\ -0,132 & -0,122 & 0,298 & 0,803 & 1,000 & 0,053 & 0,223 \\ -0,401 & 0,790 & -0,447 & -0,160 & 0,053 & 1,000 & 0,801 \\ -0,304 & 0,823 & -0,234 & 0,009 & 0,223 & 0,801 & 1,000 \end{bmatrix}$$

Dari matriks korelasi (**R**) diperoleh *eigenvalue* (λ) dan *eigenvector* (**a**) sebagai berikut: $\lambda_1 = 2,9986$; $\lambda_2 = 1,9648$; $\lambda_3 = 0,9875$; $\lambda_4 = 0,6473$; $\lambda_5 = 0,1556$; $\lambda_6 = 0,1467$; $\lambda_7 = 0,0995$

$$\mathbf{a}_1 = \begin{bmatrix} 0,23299 \\ -0,52984 \\ 0,29509 \\ 0,20946 \\ 0,07582 \\ -0,53422 \\ -0,49285 \end{bmatrix}; \mathbf{a}_2 = \begin{bmatrix} 0,20586 \\ 0,00007 \\ -0,25707 \\ -0,59285 \\ -0,67067 \\ -0,13008 \\ -0,27082 \end{bmatrix}; \mathbf{a}_3 = \begin{bmatrix} 0,76237 \\ 0,04130 \\ -0,54554 \\ 0,30473 \\ 0,12921 \\ 0,04709 \\ 0,08771 \end{bmatrix}; \mathbf{a}_4 = \begin{bmatrix} 0,51516 \\ 0,38711 \\ 0,68453 \\ -0,15415 \\ -0,05962 \\ -0,10678 \\ 0,27830 \end{bmatrix};$$

$$\mathbf{a}_5 = \begin{bmatrix} 0,17425 \\ -0,06676 \\ 0,04247 \\ -0,56061 \\ 0,60978 \\ 0,37419 \\ -0,37047 \end{bmatrix}; \mathbf{a}_6 = \begin{bmatrix} -0,13617 \\ 0,24266 \\ -0,27994 \\ -0,31192 \\ 0,39031 \\ -0,73543 \\ 0,23177 \end{bmatrix}; \mathbf{a}_7 = \begin{bmatrix} -0,08843 \\ 0,71019 \\ -0,02157 \\ 0,27651 \\ 0,00557 \\ -0,05662 \\ -0,63847 \end{bmatrix}$$

Terlihat bahwa komponen utama ke-1 dan ke-2 mempunyai *eigenvalue* yang lebih besar dari satu, sehingga dalam penelitian ini dipilihlah komponen utama ke-1 dan ke-2 untuk digunakan dalam analisis selanjutnya. Proporsi kumulatif varian yang dapat dijelaskan oleh dua komponen utama terpilih adalah

$$\frac{\sum_{j=1}^2 \lambda_j}{\sum_{j=1}^7 \lambda_j} \times 100\% = \frac{2,9985881+1,9648179}{2,9985881+1,9648179+\dots+0,0995220} \times 100\% = 70,91\%$$

Selanjutnya diperoleh persamaan untuk menghitung skor komponen utama sebagai berikut:

$$K_1 = 0,23299 Z_1 - 0,52984 Z_2 + 0,29509 Z_3 + 0,20946 Z_4 + 0,07582 Z_5 - 0,53422 Z_6 - 0,49285 Z_7$$

$$K_2 = 0,20586 Z_1 + 0,00007 Z_2 - 0,25707 Z_3 - 0,59285 Z_4 - 0,67067 Z_5 - 0,13008 Z_6 - 0,27082 Z_7$$

4.4.2 Analisis Komponen Utama Robust

Dengan menggunakan metode *C-Steps* melalui algoritma FAST-MCD diperoleh matriks kovarian (**S**) dari variabel prediktor adalah sebagai berikut:

$$\mathbf{S} = \begin{bmatrix} 0,996 & -0,008 & -0,200 & 0,112 & -0,068 & -0,362 & -0,226 \\ -0,008 & 0,857 & -0,024 & -0,165 & 0,004 & 0,555 & 0,883 \\ -0,200 & -0,024 & 0,656 & 0,216 & 0,420 & 0,005 & -0,012 \\ 0,112 & -0,165 & 0,216 & 0,577 & 0,412 & -0,045 & -0,072 \\ -0,068 & 0,004 & 0,420 & 0,412 & 0,634 & 0,081 & 0,168 \\ -0,362 & 0,555 & 0,005 & -0,045 & 0,081 & 0,684 & 0,730 \\ -0,226 & 0,883 & -0,012 & -0,072 & 0,168 & 0,730 & 1,056 \end{bmatrix}$$

Dari matriks kovarian (**S**) diperoleh *eigenvalue* (λ) dan *eigenvector* (**a**) sebagai berikut:
 $\lambda_1 = 2,4529$; $\lambda_2 = 1,3771$; $\lambda_3 = 1,0135$; $\lambda_4 = 0,3868$; $\lambda_5 = 0,1363$; $\lambda_6 = 0,0884$; $\lambda_7 = 0,0047$

$$\mathbf{a}_1 = \begin{bmatrix} 0,23588 \\ -0,53267 \\ -0,02526 \\ 0,07700 \\ -0,07898 \\ -0,48566 \\ -0,64183 \end{bmatrix}; \mathbf{a}_2 = \begin{bmatrix} 0,28010 \\ 0,17165 \\ -0,56940 \\ -0,45766 \\ -0,59756 \\ -0,02839 \\ 0,02300 \end{bmatrix}; \mathbf{a}_3 = \begin{bmatrix} 0,85059 \\ 0,26466 \\ -0,01305 \\ 0,32715 \\ 0,24759 \\ -0,09191 \\ 0,17179 \end{bmatrix}; \mathbf{a}_4 = \begin{bmatrix} 0,14532 \\ 0,28803 \\ 0,69909 \\ -0,57032 \\ -0,06839 \\ -0,26911 \\ -0,06951 \end{bmatrix};$$

$$\mathbf{a}_5 = \begin{bmatrix} 0,09342 \\ 0,09657 \\ 0,36987 \\ 0,43223 \\ -0,64197 \\ 0,42704 \\ -0,25265 \end{bmatrix}; \mathbf{a}_6 = \begin{bmatrix} -0,31203 \\ 0,39653 \\ -0,04332 \\ 0,40672 \\ -0,26328 \\ -0,69411 \\ 0,16435 \end{bmatrix}; \mathbf{a}_7 = \begin{bmatrix} -0,12327 \\ 0,60600 \\ -0,21809 \\ -0,00950 \\ 0,29878 \\ 0,13532 \\ -0,67996 \end{bmatrix}$$

Terlihat bahwa komponen utama ke-1, ke-2, dan ke-3 mempunyai *eigenvalue* yang lebih besar dari satu, sehingga dalam penelitian ini dipilihlah komponen utama ke-1, ke-2, dan ke-3 untuk digunakan dalam analisis selanjutnya. Proporsi kumulatif varian yang dapat dijelaskan oleh tiga komponen utama terpilih adalah

$$\frac{\sum_{j=1}^3 \lambda_j}{\sum_{j=1}^7 \lambda_j} \times 100\% = \frac{2,452937216+1,377066269+1,013489902}{2,452937216+1,377066269+\dots+0,004693679} \times 100\% = 88,71\%$$

Selanjutnya diperoleh persamaan untuk menghitung skor komponen utama sebagai berikut:

$$W_1 = 0,23588 Z_1 - 0,53267 Z_2 - 0,02526 Z_3 + 0,07700 Z_4 - 0,07898 Z_5 - 0,48566 Z_6 - 0,64183 Z_7$$

$$W_2 = 0,28010 Z_1 + 0,17165 Z_2 - 0,56940 Z_3 - 0,45766 Z_4 - 0,59756 Z_5 - 0,02839 Z_6 - 0,02300 Z_7$$

$$W_3 = 0,85059 Z_1 + 0,26466 Z_2 - 0,01305 Z_3 + 0,32715 Z_4 + 0,24759 Z_5 - 0,09191 Z_6 + 0,17179 Z_7$$

4.5 Regresi Komponen Utama

Regresi komponen utama dilakukan dengan meregresikan komponen-komponen utama terpilih dengan variabel respon. Apabila terdapat komponen utama yang tidak signifikan maka dikeluarkan dari model dan dilakukan analisis regresi kembali tanpa menggunakan komponen utama yang tidak signifikan.

1. Pemodelan Indeks Pembangunan Manusia di Jawa Tengah menggunakan analisis komponen utama klasik yang dikombinasikan dengan analisis regresi OLS menghasilkan persamaan model sebagai berikut:

$$\hat{Q} = -0,43071 K_1$$

$$\hat{Y} = 40047515596,88450 - 1,32624 X_1 + 10,02944 X_2 - 2,35162 X_3 + 81555,22810 X_4 - 1,70587 X_5 + 14,98813 X_6 + 7,52236 X_7$$

dengan nilai *Adjusted R*² sebesar 0,5432 dan RSE sebesar 0,6661.

Karena jumlah data (*n*) = 35 dan banyaknya parameter (*p*) = 1, maka batas suatu data dikatakan sebagai *outlier* jika nilai $|DFFITs_i| > 2\sqrt{p/n} = 0,3380617$.

Berdasarkan *output* deteksi *outlier* yang dilakukan dengan uji *DFFITs* didapatkan bahwa ada sebanyak 6 *outlier*, yaitu data ke-5, 6, 7, 31, 32, dan 33.

2. Pemodelan Indeks Pembangunan Manusia di Jawa Tengah menggunakan analisis komponen utama klasik yang dikombinasikan dengan analisis regresi robust LTS menghasilkan persamaan model sebagai berikut:

$$\hat{Y} = 76246704518,65540 + 0,12644 X_1 + 10,30716 X_2 - 3,19767 X_3 - 155273,48156 X_4 - 8,54883 X_5 + 10,09273 X_6 + 7,42005 X_7$$

dengan nilai *Adjusted R*² sebesar 0,5590 dan RSE sebesar 0,6545.

Karena jumlah data (*n*) = 35 dan banyaknya parameter (*p*) = 2, maka batas suatu data dikatakan sebagai *outlier* jika nilai $|DFFITs_i| > 2\sqrt{p/n} = 0,4780914$.

Berdasarkan *output* deteksi *outlier* yang dilakukan dengan uji *DFFITs* didapatkan bahwa ada sebanyak 5 *outlier*, yaitu data ke-2, 5, 6, 32, dan 33.

Karena terdapat *outlier* pada residual model regresi komponen utama yang diestimasi dengan metode *Ordinary Least Squares* (OLS), maka penelitian dilanjutkan dengan metode *Least Trimmed Squares* (LTS) sebagai pengganti metode OLS dalam melakukan analisis regresi komponen utama.

3. Pemodelan Indeks Pembangunan Manusia di Jawa Tengah menggunakan analisis komponen utama robust MCD yang dikombinasikan dengan analisis regresi OLS menghasilkan persamaan model sebagai berikut:

$$\hat{Q} = -0,40214 K_1$$

$$\hat{Y} = 37391070381,84630 - 1,23827 X_1 + 9,36416 X_2 - 2,19563 X_3 - 76145,47939 X_4 - 1,59271 X_5 + 13,99393 X_6 + 7,02339 X_7$$

dengan nilai *Adjusted R*² sebesar 0,6030 dan RSE sebesar 0,5523.

4. Pemodelan Indeks Pembangunan Manusia di Jawa Tengah menggunakan analisis komponen utama robust MCD yang dikombinasikan dengan analisis regresi robust LTS menghasilkan persamaan model sebagai berikut:

$$\hat{Q} = -0,26396 W_1 + 0,36136 W_2$$

$$\hat{Y} = 82435256442,11270 + 0,51482 X_1 + 8,90539 X_2 - 3,68366 X_3 - 167876,22840 X_4 - 10,19071 X_5 + 7,68224 X_6 + 5,70909 X_7$$

dengan nilai *Adjusted R²* sebesar 0,6913 dan RSE sebesar 0,4690.

5. KESIMPULAN

Berdasarkan hasil dan pembahasan, dapat diambil kesimpulan bahwa regresi komponen utama robust MCD-LTS paling efektif untuk penanganan multikolinieritas dan outlier pada pemodelan Indeks Pembangunan Manusia di Jawa Tengah. Efektivitas penerapan metode MCD dan LTS dalam pemodelan Indeks Pembangunan Manusia di Jawa Tengah ditunjukkan dengan nilai *Adjusted R²* sebesar 0,6913 dan RSE sebesar 0,4690. Nilai-nilai tersebut merupakan nilai *Adjusted R²* yang paling besar dan RSE yang paling kecil jika dibandingkan dengan ketiga model lainnya.

DAFTAR PUSTAKA

- Anisa, R. 2010. *Penggunaan Analisis Komponen Utama Kekar dan Regresi Kekar dalam Pendugaan Curah Hujan Lokal di Indramayu*. <https://s3.amazonaws.com/ppt-download/skripsirahmaanisa-120723220346-phpapp01.pdf?response-content-disposition=attachment&Signature=ryAfEglLdU289Sp9VdKVdGkWdZI%3D&Expires=1548591267&AWSAccessKeyId=AKIAIA5TS2BVP74IAVEQ>. Diakses: 4 November 2018.
- Ekosiswoyo, R., Kardoyo, dan Raharjo, T. J. 2008. *Strategi Akselerasi Pencapaian IPM Bidang Pendidikan untuk Mendukung Keberhasilan Pembangunan Jangka Menengah Kota Semarang*. Riptek Vol. 1, No. 2: Hal. 23-33.
- Filzmoser, P. 1999. *Robust Principal Component and Factor Analysis in The Geostatistical Treatment of Environmental Data*. Environmetrics Vol. 10, Hal. 363-375.
- Gujarati, D. 2004. *Ekonometrika Dasar*. Diterjemahkan oleh: Sumarno Zain. Jakarta: Erlangga. Terjemahan dari: *Basic Econometrics*.
- James, G., Witten, D., Hastie, T., dan Tibshirani, R. 2013. *An Introduction to Statistical Learning with Applications in R*. New York: Springer.
- Johnson, R.A. dan Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*. Sixth Edition. New Jersey: Pearson Prentice Hall.
- Jumiati, S., Martha, S., dan Imro'ah, N. 2018. *Penerapan Analisis Komponen Utama untuk Mereduksi Variabel dalam Pengukuran Desain Helm*. Bimaster Vol. 7, No. 3: Hal. 225-230.
- Kutner M. H., Nachtsheim, C. J., Neter, J., dan Li, W. 2005. *Applied Linear Statistical Models*. Fifth Edition. New York: The McGraw-Hill Companies, Inc.
- Marcus, G. L., Wattimanela, H. J., dan Lesnussa, Y.A. 2012. *Analisis Komponen Utama untuk Mengatasi Masalah Multikolinieritas dalam Analisa Regresi Linier Berganda (Studi Kasus: Curah Hujan di Kota Ambon Tahun 2010)*. Jurnal Berekeng Vol. 6, No. 1: Hal. 31-40.
- Montgomery, D. C., Peck, E. A., dan Vining, G. G. 2012. *Introduction to Linear Regression Analysis*. Fifth Edition. New Jersey: John Wiley & Sons, Inc.
- Notiragayu dan Nisa, K. 2008. *Analisis Regresi Komponen Utama Robust untuk Data Mengandung Pencilan*. Jurnal Sains MIPA Vol. 14, No.1: Hal. 45-50.

- Perihatini, D. I. 2018. *Perbandingan Metode Estimasi LTS, Estimasi M, dan Estimasi S pada Regresi Robust (Studi Kasus: Pembiayaan Mobil pada Perusahaan 'X' Tahun 2016)*. <https://dspace.uui.ac.id/bitstream/handle/123456789/6590/14611091%20Denisha%20Intan%20Perihatini.pdf?sequence=1&isAllowed=y>. Diakses: 6 November 2018.
- Putra, D. M. dan Ratnasari, V. 2015. *Pemodelan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur dengan Menggunakan Metode Regresi Logistik Ridge*. *Jurnal Sains dan Seni ITS* Vol. 4, No. 2.
- Rousseeuw, P. J. dan Driessen, K. V. 1999. *A Fast Algorithm for The Minimum Covariance Determinant Estimator*. *Technometrics* Vol. 41, No. 3: Hal. 212-223.
- Rousseeuw, P. J. dan Driessen, K. V. 2006. *Computing LTS Regression for Large Data Sets*. *Data Mining and Knowledge Discovery* Vol. 12: Hal. 29-45.
- Rousseeuw P. J. dan Leroy, A.M. 1987. *Robust Regression and Outlier Detection*. New York: John Wiley & Sons, Inc.
- Safidah, A. 2014. *Analisis Dekomposisi Spektral dengan Metode Principal Component Analysis*. <http://etheses.uin-malang.ac.id/6842/1/09610121.pdf>. Diakses: 27 Oktober 2018.
- Soemartini. 2008. *Principal Component Analysis (PCA) sebagai Salah Satu Metode untuk Mengatasi Masalah Multikolinieritas*. <https://elmurobbie.files.wordpress.com/2009/06/principal-component-analysis-pca2.pdf>. Diakses: 27 Oktober 2018.
- Srinadi, I G. A. M. 2014. Pengaruh Outlier terhadap Estimator Parameter Regresi dan Metode Regresi Robust. *Prosiding Konferensi Nasional Matematika XVII-2014*. Surabaya, 11-14 Juni 2014: Institut Teknologi Sepuluh November.
- Sriningsih, M., Hatidja, D., dan Prang, J. D. 2018. *Penanganan Multikolinieritas dengan Menggunakan Analisis Regresi Komponen Utama pada Kasus Impor Beras di Provinsi Sulawesi Utara*. *Jurnal Ilmiah Sains* Vol. 18, No. 1.
- Widarjono, A. 2005. *Ekonometrika: Teori dan Aplikasi untuk Ekonomi dan Bisnis*. Edisi Pertama. Yogyakarta: Ekonisia.
- Widiharih, T. 2001. *Penanganan Multikolinieritas (Kekolinieran Ganda) dengan Analisis Regresi Komponen Utama*. *Jurnal Matematika dan Komputer* Vol. 4, No. 2: Hal. 71-81.
- Willems, G. dan Aelst, S. V. 2005. *Fast and Robust Bootstrap for LTS*. *Computational Statistics and Data Analysis* Vol. 48: Hal. 703-715.