

**PEMODELAN PERSENTASE KANDUNGAN LEMAK TUBUH  
DENGAN *PRINCIPAL COPONENT REGRESSION* (PCR)  
DAN *PARTIAL LEAST SQUARE REGRESSION* (PLSR)**

***MODELING FOR BODY FAT PERCENTAGE ESTIMATES WITH  
PRINCIPAL COPONENT REGRESSION (PCR) AND PARTIAL  
LEAST SQUARE REGRESSION (PLSR)***

**Hedi**

Jurusan Teknik Konversi Energi Politeknik Negeri Bandung  
hedi@polban.ac.id

**ABSTRAK**

Kandungan lemak tubuh, jika bisa diprediksi tinggi atau rendah, bisa dijadikan referensi untuk melihat tingkat kesehatan seseorang. Semakin tinggi kandungan lemak, semakin besar risiko berbagai penyakit. Penelitian ini akan mengusulkan pemodelan kandungan lemak sebagai respons yang bergantung pada tiga belas variabel prediktor yaitu: usia, berat, tinggi, leher, dada, perut, pinggul, paha, lutut, pergelangan kaki, bisep, lengan, dan pergelangan tangan. Masalah yang dihadapi dalam model ini adalah bahwa antarprediktor berkorelasi (multikolinieritas). Untuk itu, diterapkan metode PCR dan PLSR. Tujuan penelitian ini adalah untuk memodelkan persentase lemak tubuh. Dengan menerapkan MSPE, model PLSR adalah yang paling baik untuk memodelkan prediksi persentase kandungan lemak tubuh.

**Kata Kunci :** Lemak Tubuh, Pemodelan, Multikolinieritas, PCR, PLSR

***ABSTRACT***

*If the body fat content can be predicted high or low, it can be used as a reference to see the level of health. The higher the fat content, the greater the risk of various diseases. This paper proposed a modeling fat content as a response that depends on 13 predictor variables; they are age, weight, height, neck, chest, abdomen, hips, thighs, knees, ankles, biceps, arms, and wrists. The problem faced in this model is that the predictor variables are correlated (multicollinearity). The objective of this paper is to determine a fit model for predicted body fat percentage. To overcome these problems, PCR and PLSR methods were applied. By applying MSPE, the PLSR model is the best for modeling the prediction of the body fat percentage.*

**Keywords:** *Body Fat, Modelling, Multicollinearity, PCR, PLSR*

**PENDAHULUAN**

Tubuh perlu lemak untuk mengatur produksi hormon serta membantu penyerapan beberapa jenis vitamin. Kadar lemak tubuh normal dan sehat adalah wanita 10 - 31% dan pria 2 - 24%. Kandungan lemak tubuh di atas 32% untuk wanita dan di atas 25% pria dikategorikan gemuk

(Kiddie, 2018). Beberapa metode untuk mengukur kandungan lemak tubuh adalah dengan menggunakan alat ukur kandungan lemak tubuh, dengan teknik penimbangan bawah air, dan penimbangan berat di udara, selisih berat menghasilkan kandungan (Roger, 1995). Pemodelan kandungan lemak tubuh bergantung pada tiga belas ukuran, yaitu berat badan, tinggi badan, lingkar

leher, lingkaran dada, lingkaran perut, lingkaran pinggul, lingkaran paha, lingkaran lutut, lingkaran pergelangan kaki, lingkaran bicep, lingkaran lengan, dan lingkaran pergelangan tangan. Ukuran ini dapat dilakukan dengan model *multiple regression*. Model ini mengasumsikan tidak terdapat korelasi di antara prediktor yang dalam kenyataannya, terdapat korelasi yang cukup signifikan di antara prediktor, yaitu berat badan berkorelasi dengan tinggi badan. Artinya, semakin tinggi seseorang, semakin berat orang tersebut. Berat dapat pula berkorelasi dengan lingkaran perut, lingkaran lutut, lingkaran pergelangan kaki, lingkaran bicep, lingkaran lengan, dan lingkaran pergelangan tangan. Lingkaran perut berkorelasi dengan lingkaran paha dan sebagainya (masalah multikolinieritas). Masalah ini akan berimplikasi pada estimasi koefisien regresi dengan *standard error* sangat tinggi sehingga tidak dapat menjelaskan dengan baik hubungan antara variabel respon dan variabel-variabel prediktor (Jamal, 2017). Model persamaan lemak tubuh dapat disederhanakan dengan pendekatan metode *multiple regression* melalui penyederhanaan prediktor menjadi berat, tinggi, pinggul, dan pinggang (Lorenzo *et al*, 2014). Kelemahan metode ini adalah prediktor lainnya, seperti ukuran lingkaran perut, lingkaran leher, dan usia, tidak diperhatikan.

Beberapa metode yang diperkenalkan untuk menanggulangi masalah multikolinearitas adalah *Principal Component Regression* (PCR) dan *Partial least square Regression* (PLSR) (Hall, 2014). Dalam penelitian ini, ditentukan model persentase kandungan lemak tubuh dengan PCR dan PLSR

### **Multiple Regression**

Analisis *Multiple Regression* adalah teknik analisis statistik yang menjelaskan hubungan di antara variabel respon Y dan variabel prediktor  $X_1, X_2, X_3, \dots, X_p$ . Dengan

menerapkan data sampel yang merupakan pasangan prediktor dan *respon*, taksiran parameter model ditentukan dengan metode *Ordinary Least Squares* (OLS) (Walpole, 2012). Permasalahan yang sering ditemukan dalam analisis *multiple regression* adalah adanya korelasi yang tinggi antara variabel prediktor yang berimplikasi pada solusi *least square* tidak unik (Gujarati, D.N., 2009).

### **PCR**

Salah satu metode untuk menanggulangi masalah multikolenieritas adalah PCR. Melalui metode ini, variabel bebas  $X_1, X_2, X_3, \dots, X_p$  yang saling berkorelasi (multikolinearitas) ditransformasi ke dalam variabel komponen  $C_1, C_2, C_3, \dots, C_k$  yang tidak saling berkorelasi. Kombinasi linier yang dipilih adalah kombinasi linier dengan variansi  $V(C_1) > V(C_2) > V(C_3) > \dots > V(C_k)$  dan korelasi variabel antarkomponen  $C_1, C_2, C_3, \dots, C_k$  adalah nol. Regresi respon Y dengan komponen  $C_1, C_2, C_3, \dots, C_k$  menghasilkan persamaan Y dengan k buah komponen. Dengan menyubstitusikan persamaan komponen dengan variabel prediktor, didapat persamaan model respon dengan variabel prediktor (Hastie *et al*, 2009).

### **PLSR**

Regresi respon Y dengan komponen  $C_1, C_2, C_3, \dots, C_k$  dilakukan dengan transformasi variabel prediktor yang melibatkan variabel respon, yaitu mendekomposisi matriks variabel prediktor menjadi matriks komponen dan matriks *loading*. Matriks komponen dipilih sedemikian rupa sehingga komponen yang terbentuk dapat menjelaskan variasi pada variabel prediktor dan variabel respon melalui fungsi kovariansi sampel antara setiap variabel prediktor dengan variabel respon. Hal tersebut dibutuhkan agar komponen yang terbentuk tidak hanya baik untuk menerangkan variasi dari variabel-variabel

prediktor, tetapi juga baik dan relevan untuk memprediksi variabel respon (Abdi, 2010). Beberapa peneliti membandingkan pendekatan model regresi *multivariate* melalui metode *Stepwise Multiple Linear Regresi* (SMLR) dengan Metode PLSR dalam menentukan total kandungan nitrogen di tanah. Hasilnya adalah metode PLSR yang paling baik (Shi, et al 2013). Selanjutnya, model PCR dan PLSR yang diterapkan untuk memprediksi kadar garam tanah menggunakan *soil-reflected spectra*. Hasilnya adalah model berdasarkan semua pita spektral asli menunjukkan bahwa PLSR lebih unggul daripada PCR (Zhang et al, 2019).

## METODE

Berdasarkan respons lemak Y dan predictor  $X_1, X_2, X_3, \dots, X_p$  yang saling berkorelasi (multikolinearitas), dibuat matriks data prediktor  $X_{n \times p}$  karena satuan pengukuran data predictor tidak sama sehingga data dibakukan dalam bentuk matriks  $Z_{n \times p}$ . Selanjutnya, variabel  $Z_1, Z_2, Z_3, \dots, Z_p$  yang saling berkorelasi ditransformasikan ke dalam variabel komponen  $C_1, C_2, C_3, \dots, C_k$  yang tidak saling berkorelasi.

$$C_i = a_{1i}Z_1 + a_{2i}Z_2 + a_{3i}Z_3 \dots a_{ji}Z_p \quad (1)$$

dengan  $i = 1, 2, 3, \dots, k$  ;  $j = 1, 2, 3, \dots, p$

Kombinasi linier yang dipilih adalah kombinasi linier dengan variansi  $V(C_1) > V(C_2) > V(C_3) > \dots > V(C_k)$  dan korelasi variabel antarkomponen  $C_1, C_2, C_3, \dots, C_k$  adalah nol. Berdasarkan regresi respon  $\hat{Y} = Y - \bar{Y}$  dengan k variabel komponen, didapat taksiran

$$\hat{\bar{Y}} = b_0 + b_1C_1 + b_2C_3 + \dots + b_kC_k \quad (2)$$

Selanjutnya, substitusi persamaan satu ke persamaan dua didapat taksiran respon Y dengan data prediktor  $X_1, X_2, X_3, \dots, X_p$  sebagai berikut.

$$\hat{Y} = b_0^* + b_1^*X_1 + b_2^*X_2 + b_3^*X_3 + \dots + b_p^*X_p \quad (3)$$

Dengan metode PLSR, matriks data prediktor X dan respon Y dibakukan kemudian matriks variabel prediktor didekomposisi menjadi matriks komponen *orthogonal* T dan matriks *loading* P. Matriks komponen dipilih sedemikian rupa sehingga komponen yang terbentuk dapat menjelaskan variasi pada variabel prediktor dan variabel respon melalui fungsi kovariansi sampel antara setiap variabel prediktor dengan variabel respon. Hal tersebut dibutuhkan agar komponen yang terbentuk tidak hanya baik dalam menerangkan variasi dari variabel-variabel prediktor, tetapi juga baik dan relevan dalam memprediksi variabel respon. Dalam PLS, komponen-komponen tersebut akan berperan sama seperti variabel prediktor kemudian dilakukan seleksi komponen menggunakan algoritma NIPALS (Hedi, 2017).

Untuk memilih model kandungan lemak tubuh dari kedua metode, ditentukan melalui perhitungan MSPE (*mean squared prediction error*).

$$MSPE = \frac{1}{n} \sum_{t=1}^h (y_t - \hat{y}_t)^2 \quad (4)$$

Metode yang paling cocok adalah yang memberikan nilai MSPE terkecil dari kedua metode hitungan tersebut .

## HASIL DAN PEMBAHASAN

Data kandungan lemak tubuh yang bergantung pada tiga belas variabel prediktor adalah data sekunder, dengan respon Y adalah kandungan lemak tubuh yang bergantung pada tiga belas variabel *predictor* yaitu  $X_1$ : *Age* dalam tahun,  $X_2$  : *Weight* dalam lbs,  $X_3$  : *Height* dalam inci,  $X_4$  : *Neck circumference* dalam satuan cm,  $X_5$  : *Chest circumference* dalam satuan cm,  $X_6$  : *Abdomen 2 circumference* dalam satuan cm,  $X_7$  : *Hip circumference* dalam satuan cm,  $X_8$  : *Thigh circumference* dalam satuan cm,  $X_9$  : *Knee circumference* dalam satuan cm,  $X_{10}$  : *Ankle circumference* dalam satuan cm,  $X_{11}$  :

*Biceps (extended) circumference* dalam satuan cm,  $X_{12}$ : *Forearm circumference* dalam satuan cm, dan  $X_{13}$  : *Wrist circumference* dalam satuan cm (Roger W. Johnson, 1995). Kandungan lemak minimum dari 252 yang diamati adalah

minimum 0% dan maksimum 47,5%, dengan rata-rata 19% dan simpangan baku 8,3. Usia yang diamati minimum 22 tahun dan maksimum 81 tahun dengan umur rata-rata adalah 45 tahun (tabel 1).

**Tabel 1. Deskripsi Data**

| Variable     | Observations | Minimum | Maximum | Mean    | Std. deviation |
|--------------|--------------|---------|---------|---------|----------------|
| Percent      | 252          | 0       | 47.5    | 19.151  | 8.369          |
| Age          | 252          | 22      | 81      | 44.885  | 12.602         |
| Weight       | 252          | 118.5   | 363.15  | 178.924 | 29.389         |
| Height       | 252          | 29.5    | 77.75   | 70.149  | 3.663          |
| Neck CF      | 252          | 31.1    | 51.2    | 37.992  | 2.431          |
| Chest CF     | 252          | 79.3    | 136.2   | 100.824 | 8.43           |
| Abdomen 2 CF | 252          | 69.4    | 148.1   | 92.556  | 10.783         |
| Hip CF       | 252          | 85      | 147.7   | 99.905  | 7.164          |
| Thigh CF     | 252          | 47.2    | 87.3    | 59.406  | 5.25           |
| Knee CF      | 252          | 33      | 49.1    | 38.59   | 2.412          |
| Ankle CF     | 252          | 19.1    | 33.9    | 23.102  | 1.695          |
| Biceps CF    | 252          | 24.8    | 45      | 32.273  | 3.021          |
| Forearm CF   | 252          | 21      | 34.9    | 28.664  | 2.021          |
| Wrist CF     | 252          | 15.8    | 21.4    | 18.23   | 0.934          |

### PCR dan PLSR

Berdasarkan data sampel respon ukuran 252 dan prediktor ukuran 252x13, dibangun matrik data prediktor ukuran 252 x 13. Karena satuan pengukuran berbeda, setiap kolom pada matriks tersebut dibakukan, yaitu *mean* = 0 dan standar deviasi 1. Langkah

berikutnya adalah mentransformasikan variabel prediktor ke dalam variabel komponen yang tidak saling berkorelasi sehingga diperoleh tiga komponen. Melalui metode PCR, ditentukan regresi antara kandungan lemak tubuh dengan tiga komponen  $C_1$ ,  $C_2$ ,  $C_3$  dengan hasil perhitungannya dapat dilihat di Tabel 2.

**Tabel 2. Regresi Model PCR**

| Regression Summary for Dependent Variable: Y (Spreadsheet3) R= .75619613 R <sup>2</sup> = .57183259 Adjusted R <sup>2</sup> = .56665314 F(3,248)=110.40 p |           |                    |          |                 |          |          |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|--------------------|----------|-----------------|----------|----------|
|                                                                                                                                                           | Beta      | Std.Err. - of Beta | B        | Std.Err. - of B | t(248)   | p-level  |
| Intercept                                                                                                                                                 |           |                    | 19.04159 | 0.355952        | 53.49481 | 0.000000 |
| C1                                                                                                                                                        | 0.615127  | 0.041551           | 1.86211  | 0.125784        | 14.80407 | 0.000000 |
| C2                                                                                                                                                        | 0.395870  | 0.041552           | 2.91920  | 0.306411        | 9.52709  | 0.000000 |
| C3                                                                                                                                                        | -0.193520 | 0.041552           | -1.65909 | 0.356233        | -4.65733 | 0.000005 |

Persamaan kandungan lemak tubuh  
 $\hat{Y} = 19,04 + 1,86 C_1 + 2,92 C_2 - 1,66 C_3$  (5)  
 $Adjusted R^2 = 56,7\%$ , MSPE = 22,71 dan  
pengaruh variabel independen  $C_1$ ,  $C_2$ ,  $C_3$   
secara bersama-sama terhadap respon  
kandungan lemak tubuh dilakukan  
pengujian hipotesis, dengan bentuk uji  
hipotesis nol ( $H_0$ ) seluruh koefisien regresi

$\beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$  dan hipotesis alternatif  
( $H_1$ ) sehingga terdapat satu  $\beta_i = 0$  untuk  $i =$   
0, 1, 2, 3. Dengan kata lain,

$H_0$ : tidak ada pengaruh variabel komponen  
terhadap respon kandungan lemak tubuh

$H_1$ : terdapat pengaruh variabel komponen  
terhadap respon kandungan lemak tubuh

**Tabel 3. Analisis Variansi Model PCR**

| Analysis of Variance; DV: Var4 (Spreadsheet2) |                   |     |                |          |          |
|-----------------------------------------------|-------------------|-----|----------------|----------|----------|
|                                               | Sums of - Squares | df  | Mean - Squares | F        | p-level  |
| <b>Regress.</b>                               | 10575.25          | 3   | 3525.082       | 110.4042 | 0.000000 |
| <b>Residual</b>                               | 7918.36           | 248 | 31.929         |          |          |
| <b>Total</b>                                  | 18493.61          |     |                |          |          |

Berdasarkan tabel 3, didapat p-  
level = 0,000 yang berarti  $H_0$  ditolak.  
Artinya, ada pengaruh secara signifikan  
variabel komponen terhadap respon

kandungan lemak. Selanjutnya, melalui  
metode PLSR regresi kandungan lemak  
tubuh dengan tiga komponen  $C_1$ ,  $C_2$ ,  $C_3$ ,  
diperoleh hasil seperti dilihat di Tabel 4.

**Tabel 4. Regresi Model PLSR**

| Regression Summary for Dependent Variable: Var4 (PCR3) R= .82226019 R <sup>2</sup> = .67611182 Adjusted R <sup>2</sup> = .67219382 F(3,248)=172.57 p |          |                    |          |                 |          |          |
|------------------------------------------------------------------------------------------------------------------------------------------------------|----------|--------------------|----------|-----------------|----------|----------|
|                                                                                                                                                      | Beta     | Std.Err. - of Beta | B        | Std.Err. - of B | t(248)   | p-level  |
| <b>Intercept</b>                                                                                                                                     |          |                    | 19.04159 | 0.309587        | 61.50649 | 0.000000 |
| <b>C1</b>                                                                                                                                            | 0.667276 | 0.036139           | 2.09862  | 0.113658        | 18.46435 | 0.000000 |
| <b>C2</b>                                                                                                                                            | 0.429011 | 0.036139           | 3.32078  | 0.279733        | 11.87126 | 0.000000 |
| <b>C3</b>                                                                                                                                            | 0.216342 | 0.036139           | 2.83513  | 0.473593        | 5.98644  | 0.000000 |

Persamaan kandungan lemak  
tubuh metode PLSR  
 $\hat{Y} = 19,04 + 2,10 C_1 + 3,32 C_2 - 2,83 C_3$  (6)  
 $Adjusted R^2 = 67,2\%$ , MSPE = 21,72,  
dan pengaruh variabel independen  $C_1$ ,

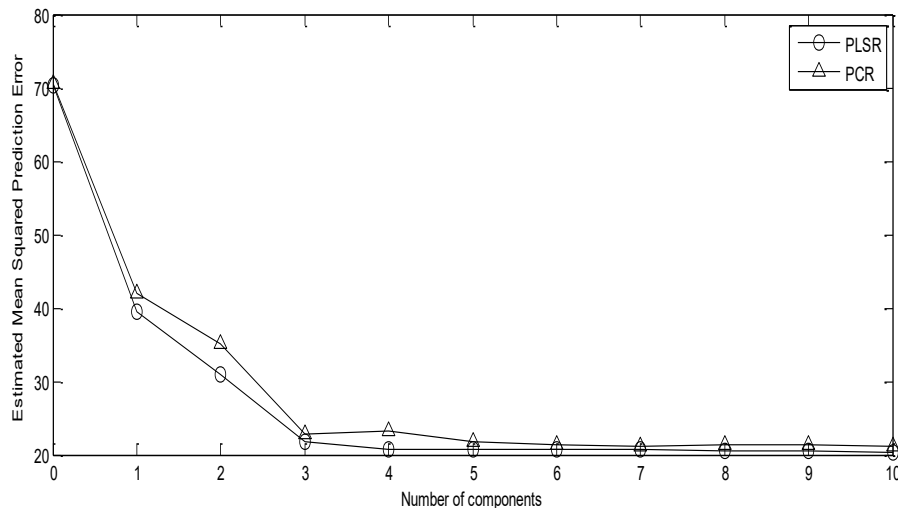
$C_2$ ,  $C_3$  secara bersama-sama terhadap  
respon kandungan lemak tubuh adalah  
signifikan seperti dapat dilihat pada  
Tabel 5.

Tabel 5. Analisis Variansi Model PLSR

| Analysis of Variance; DV: Var4 (PCR3) |                   |     |                |          |         |
|---------------------------------------|-------------------|-----|----------------|----------|---------|
|                                       | Sums of - Squares | df  | Mean - Squares | F        | p-level |
| <b>Regress.</b>                       | 12503.75          | 3   | 4167.915       | 172.5655 | 0.00    |
| <b>Residual</b>                       | 5989.86           | 248 | 24.153         |          |         |
| <b>Total</b>                          | 18493.61          |     |                |          |         |

Berdasarkan hasil perhitungan MSPE model PLSR dan PCR untuk setiap banyaknya komponen yang masuk dalam model, ditunjukkan pada **Gambar 1**. Tampak pada gambar tersebut, semua variabel komponen yang masuk dalam

model nilai MSPE model PLSR berada di bawah PCR. Semakin banyak komponen yang masuk ke dalam model PCR, nilai MSPE model PCR mendekati model PLSR.

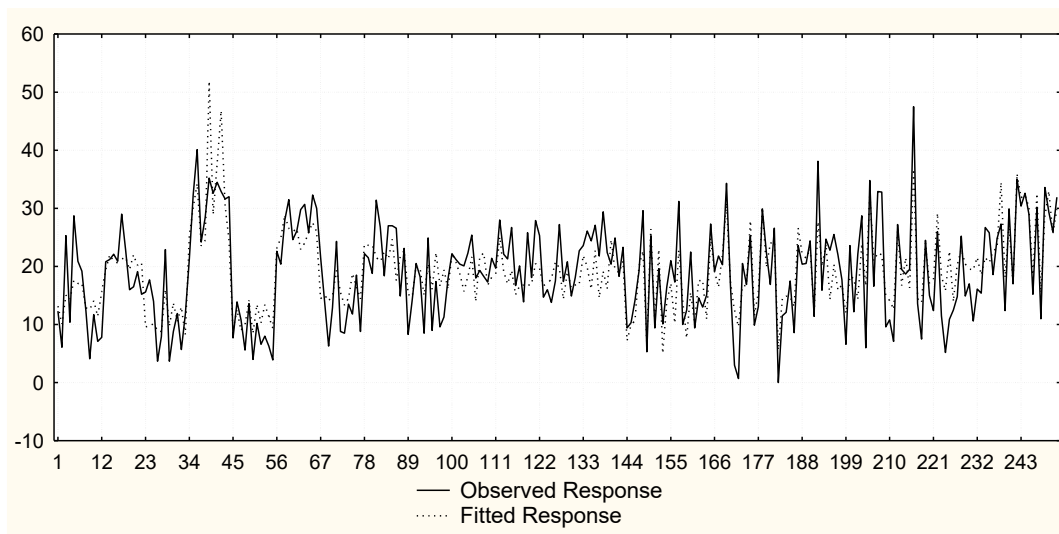


Gambar 1. Plot nilai MSPE Model PLSR dan PCR

Model prediksi kandungan lemak tubuh yang bergantung pada tiga belas prediktor melalui PLSR adalah

$$\begin{aligned} \text{Percent} = & -11,30885 + 0,15576 \\ & \text{Age} + 0,02283 \quad \text{Weight} - 0,46484 \\ & \text{Height} + 0,06238 \quad \text{Neck CF} + 0,19616 \end{aligned}$$

$$\begin{aligned} & \text{Chest CF} + 0,23979 \quad \text{Abdomen} \quad 2 \\ & \text{CF} + 0,13559 \quad \text{Hip CF} + 0,13117 \quad \text{Thigh CF} \\ & + 0,12580 \quad \text{Knee CF} - 0,34391 \quad \text{Ankle} \\ & \text{CF} + 0,09687 \quad \text{Biceps CF} - 0,04433 \\ & \text{Forearm CF} - 0,68318 \quad \text{Wrist CF} \quad (7) \end{aligned}$$



Gambar 2. Plot Kesesuaian Model PLSR

Persentase lemak tubuh dari data pengamatan dan hasil prediksi kandungan lemak diperoleh melalui model persamaan 7. Prediksi persentase lemak ditandai sebagai *fitted Response* model PLSR dan persentase kandungan lemak ditandai dengan *observed response*. Terlihat bahwa prediksi lemak tubuh sesuai dengan nilai data respon.

## SIMPULAN

Metode PCR dan PLSR yang diterapkan dalam memodelkan persentase kandungan lemak, model PLSR yang terbaik dengan nilai MSPE = 21,72 untuk model PLSR dan nilai MSPE = 22,71 untuk model PCR. Melalui perhitungan *Adjusted R<sup>2</sup>*, model PCR adalah 56,7% dan melalui model PLSR, terjadi peningkatan dengan *Adjusted R<sup>2</sup>* = 67,2%. Model persamaan kandungan lemak yang bergantung pada tiga belas prediktor sebagai berikut.

$$\text{Percent} = -11,30885 + 0,15576 \text{ Age} + 0,02283 \text{ Weight} - 0,46484 \text{ Height} + 0,06238 \text{ Neck CF} + 0,19616 \text{ Chest CF} + 0,23979 \text{ Abdomen 2 CF} + 0,13559 \text{ Hip CF} + 0,13117 \text{ Thigh CF} + 0,12580 \text{ Knee CF} - 0,34391 \text{ Ankle CF} + 0,09687 \text{ Biceps CF} - 0,04433 \text{ Forearm CF} - 0,68318 \text{ Wrist CF}.$$

## DAFTAR PUSTAKA

- Abdi, Herve. 2010. *Partial Least Squares Regression and Projection on latent Structure Regression (PLS Regression)*. John Wiley & Sons, Inc.
- Daoud, Jamal I. 2017. "Multicollinearity and Regression Analysis", *Journal of Physics Conference Series* 949(1):012009, December.
- De Lorenzo; Nardi ; Iacopino ; Domino ; Murdolo; Gavrilu ; Minella ; Scapagnini ; Di Renzo. 2014. "A New Predictive Equation for Evaluating Women Body Fat Percentage and Obesity-Related Cardiovascular Disease Risk", *J Endocrinol Invest DOI*, Springer.
- Gujarati, D. N. 2009. *Basic Econometrics, Fourth Edition*. New York: Mc Graw-Hill.
- Hall, Simon. 2014. *Implementation and Verification of a Robust PLS Regression Algorithm*. Master's (Thesis in Engineering Mathematics and Computational Science). Sweden: Department of Mathematical Sciences Mathematical Statistics Chalmers, University of Technology Gothenburg.

- Hastie, Trevor et al. 2009. "The Elements Of Statistical Learning Data Mining, Inference, and Prediction". Springer.
- Hedi. 2017. "Penaksiran Parameter PLS dengan Metode Algoritma NIPALS Menggunakan Bahasa Pemrograman STATISTICA", *SIGMA-Mu* Vol.9 No.2 –2017
- Johnson, Roger W. 1995. "StatLib-Datasets Archive".  
Lib.stat.cmu.edu/datasets/Under  
filename bodyfat.  
<http://lib.stat.cmu.edu/datasets/bodyfat>  
2019
- Julian, Inzenman Alan. 2008. "Modern Multivariate Statistical Techniques" Springer
- Kiddie, Joy. 2018. "A New Estimator of Whole Body Fat Percentage".  
<https://www.lchfrd.com/2018/08/28/a-new-estimator-of-whole-body-fat-percentage/>
- Roger, Johnson. 1995. "A Multiple Regression Project", *Teaching Statistics an International Journal for Teacher* Volume 17, Issue 2.
- Shi, Tiezhu et al. 2013. "Comparison of Multivariate Methods for Estimating Soil Total Nitrogen with Visible/Near-Infrared Spectroscopy", *Plant and Soil*. Dordrecht Vol. 366, Iss. 1-2,
- Walpole, Ronald E. 2012. *Probability & Statistics for engineer & Science*. Prentice Hall
- Zhang, Xiaoguang; Huang, Biao. 2019. "Prediction of Soil Salinity with Soil-Reflected Spectra: A Comparison of Two Regression Methods", *Scientific Reports* (NaturePublisher Group), (Mar 2019): 1-8. Vol. 9. London