Aplikasi Metode *Weighted Principal Component Analysis* (WPCA) dengan Software S-PLUS2000

DEWI RACHMATIN
Departemen Pendidikan Matematika FPMIPA UPI dewirachmatin@upi.edu

Intisari: Sebuah metode baru untuk mereduksi ruang berdimensi tinggi yang dikembangkan dari metode PCA yaitu *Weighted Principal Component Analysis* (WPCA) diperkenalkan oleh J.F. Pinto da Costa, H. Alonso dan L. Roque (2011). Oleh karena pada metode PCA, koefisien korelasi Pearson sangat sensitif dengan kehadiran gangguan dan pencilan, maka pada metode WPCA ini digunakan koefisien korelasi baru yang melibatkan rank dari setiap pengamatan untuk setiap variabel. Untuk memberikan gambaran tentang metode WPCA ini, Pada artikel ini program untuk WPCA dibuat dengan software S-PLUS 2000 diterapkan pada data bayi (Damayanti,2008) serta data catatan waktu pelari (Johnson, 2007), dan hasilnya dibandingkan dengan hasil metode PCA klasik.

Kata kunci: *rank, koefisien korelasi, PCA dan WPCA.*

Abstract: A new method to reduce the high-dimensional space that is developed from the PCA method, namely Weighted Principal Component Analysis (WPCA) introduced by JF Pinto da Costa, H. and L. Roque Alonso (2011). Therefore, the PCA method, the Pearson correlation coefficient is very sensitive to the presence of interference an outliers, then the method used WPCA new correlation coefficient involving the rank of each observation for each variable. To give an idea of this WPCA method, In this article the program to WPCA made with software S-PLUS 2000 applied to the data babies (Damayanti, 2008) as well as the data record time runner (Johnson, 2007), and the results are compared with the results of classical PCA method.

Keywords: rank correlation coefficient, PCA and WPCA.

1 PENDAHULUAN

Principal Component Analysis (PCA) atau Analisis Komponen Utama secara luas digunakan untuk menganalisa data multivariat berdimensi tinggi, serta PCA menerangkan struktur kovarian data melalui kombinasi linier dari data asli (Johnson, 2007). Analisis Komponen utama (PCA) adalah teknik pereduksian dimensi yang sangat populer (Joliffe, 1986), yang sering diterapkan dalam banyak cabang ilmu seperti sains, psikologi, teknik, geologi, visi komputer, dan banyak cabang lainnya. PCA mereduksi jumlah variabel (p) menjadi k < p, seringkali k < 10, sehingga analisis data dapat dilakukan pada ruang yang berdimensi k. Untuk kepentingan interpretasi data lebih baik jika k ≤ 3, supaya data dapat dinterpretasikan dengan baik.

Metode PCA klasik membentuk sebuah himpunan variabel baru yang saling ortogonal, di mana variabel-variabel baru ini (disebut Komponen-Komponen utama) merupakan kombinasi linear dari variabel-variabel asal yang terpusatkan : $\tilde{X}_j = X_j - \bar{X}_j$ dengan j=1,2,…,p. Loading-loading (koefisien-koefisien) dari Komponen-Komponen utama berkorespondensi dengan vektor-vektor eigen dari matriks kovarians sampel ataupun matriks korelasi sampel. Untuk setiap vektor loading $\tilde{P}_j$, maka nilai eigen $\tilde{l}_j$ dari S menerangkan berapa banyak variabilitas data yang dijelaskan oleh $\tilde{P}_j$ melalui hubungan $Var(\tilde{P}_j) = \tilde{l}_j$. Biasanya vektor loading ini diurutkan sesuai dengan urutan menurun dari nilai eigen. Oleh karena itu, k Komponen utama yang pertama menjelaskan sebagian besar variabilitas data. Setelah memilih k, kita dapat memproyeksikan data berdimensi p pada ruang berdimensi k.

Untuk memilih jumlah yang sesuai (k) terdapat banyak kriteria. Yang sangat populer yang berdasarkan pada *scree plot* yang memaparkan nilai eigen dalam urutan menurun, sedangkan yang lain berdasarkan pada sebuah kriteria yang lebih formal menganggap variasi total yang dijelaskan oleh k Komponen utama pertama ≥ 80%.

Pada PCA klasik, pereduksian dimensi data dapat dilakukan atas matriks kovarians sampel atau matriks korelasi sampel. Pada beberapa studi kasus, penerapan metode PCA klasik tidak direkomendasikan dikarenakan PCA klasik memberikan bobot yang sama pada semua data pengamatan dan sensitif terhadap adanya gangguan dan pencilan. Dari hasil penelitian yang dilakukan oleh da Costa dkk (2011), da Costa dkk menyarankan PCA baru untuk menyelesaikan masalah-masalah yang terjadi pada PCA klasik, dan untuk menangani pencilan dan gangguan, sebaiknya gunakan rank (rangking) dari pada data asli.

Biasanya rangking yang tertinggi (rank=1) diberikan pada nilai pengamatan yang terkecil, akan tetapi hal tersebut bergantung permasalahan yang muncul (kepentingan masing-masing permasalahan). Sebagai contoh dalam penelitian da Costa pada *microarray data* (data gen untuk beberapa penyakit tumor dan kanker), bobot (rank) yang lebih tinggi diberikan pada nilai pengamatan (ekspresi gen mutlak) yang terbesar di dalam setiap variabel (gen). Hasil penelitian da Costa dkk. (2011) merekomendasikan penggunaan koefisien korelasi baru untuk rank yang diboboti dari pada koefisien korelasi Pearson yang biasa. Inilah yang akan disebut WPCA.

Pada artikel ini akan diperkenalkan metode WPCA pada bagian dua, sedangkan program untuk metode PCA klasik dan metode WPCA pada bagian tiga dibuat oleh penulis berdasarkan algoritma yang direkomendasikan oleh da Costa dkk (2011). Untuk selanjutnya program tersebut diterapkan pada dua data, yaitu data status gizi ibu hamil dan karakteristik bayi baru lahir (Damayanti, 2008), serta data catatan waktu pelari (Johnson, 2007). Hasil PCA klasik yang diperoleh dibandingkan dengan hasil WPCA untuk dua data kasus tersebut.

2 METODE *WEIGHTED PRINCIPAL COMPONENT ANALYSIS* (WPCA)

Pada PCA klasik, vektor-vektor eigen dari matriks kovarian atau matriks korelasi Pearson memuat koefisien-koefisien dari kombinasi linier dari matriks asli yang berkorespondensi dengan variabel-variabel baru (Komponen-Komponen utama). Seperti telah diketahui koefisien korelasi Pearson sangat sensitif terhadap kehadiran pencilan dan gangguan (Hubert,2004). Untuk menangani hal ini, akan digunakan rank (rangking) dari setiap pengamatan. Jadi mulailah dengan merangking pengamatan untuk setiap variabel dari satu (rangking tertinggi) sampai n (rangking terendah).

Misalkan diberikan data bivariat (X_i, Y_i), i=1,2, …,n, beri nama R_i untuk rangking variabel X_i dan Q_i untuk rangking variabel Y_i. Jika kita menghitung koefisien korelasi Pearson untuk rangking data, maka yang diperoleh adalah koefisien korelasi rank Spearman (r_S), yaitu :

$$r_S = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(Q_i - \bar{Q})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2 \sum_{i=1}^{n}(Q_i - \bar{Q})^2}}$$

di mana $\bar{R}$ dan $\bar{Q}$ adalah rata-rata rank. Jika tidak ada rank yang sama, maka untuk keperluan komputasi digunakan rumus : $r_S = 1 - \frac{6\sum_{i=1}^{n}(R_i-Q_i)^2}{n^3-n}$. Dalam hal ini jarak antara dua rank dalam rank Spearman diberikan oleh $D_i^2 = (R_i - Q_i)^2$.

Da Costa dan Soares (2005) memberikan alternatif jarak antara dua rank sebagai : $WD_i^2 = (R_i - Q_i)^2[(n - R_i + 1) + (n - Q_i + 1)] = D_i^2(2n + 2 - Ri - Qi$. Suku keduanya yaitu 2n + 2 - Ri - Qi disebut fungsi bobot linier. Akibatnya koefisien korelasi rank yang diboboti menjadi :

$$r_W = 1 - \frac{6\sum_{i=1}^{n}(R_i-Q_i)(2n+2-R_i-Q_i)}{n^4+n^3-n^2-n} \quad (1)$$

yang mempunyai nilai antara -1 dan +1. Beberapa sifat distribusi dari r_w ini dapat dilihat pada da Costa dkk (2006).

Kelemahan dari r_w ini adalah tidak dapat digunakan jika ada rank yang sama, sehingga da Costa dkk (2011) merekomendasikan ukuran jarak berikut : $W_2D_i^2 = (R_i - Q_i)^2[2n + 2 - R_i - Q_i]^2$ yang memberikan bobot yang tinggi untuk rank yang berada di urutan teratas.

Hal biasa mendefinisikan koefisien korelasi rank semacam rank Spearman sebagai fungsi linier dari jarak antara dua vektor rank. Untuk kasus kita, hal ini berkorespondensi dengan mendefinisikan koefisien dalam bentuk :

$$A + B\sum_{i=1}^{n}(R_i - Q_i)^2(2n + 2 - R_i - Q_i)^2 \quad (2)$$

di mana A dan B konstanta yang mempunyai nilai antara -1 dan 1.

Untuk menentukan A dan B, mulailah dengan mentransformasi data dan menghitung koefisien korelasi Pearson dari tansformasi data tersebut. Ekspresi yang diperoleh secara eksak berasal dari (2). Transformasi memuat penggantian nilai pengamatan ke i ke dalam nilai $R_i^{'} = R_i(2n + 2 - R_i)$. Kita lakukan hal yang sama untuk $Q_i^{'} = Q_i(2n + 2 - Q_i)$. Dapat dengan mudah dibuktikan bahwa untuk kasus tidak ada rank yang sama, nilai rata-rata berikut adalah benar :

$$\overline{R'} = \frac{1}{n}\sum_{i=1}^{n} R_i\,(2n + 2 - R_i) = \frac{(n + 1)(4n + 5)}{6}$$

dan $\overline{Q'} = \frac{1}{n}\sum_{i=1}^{n} Q_i\,(2n + 2 - Q_i) = \frac{(n+1)(4n+5)}{6}$.

Sehingga koefisien korelasi Pearson untuk transformasi ini adalah:

$$r_{W2} = \frac{\sum_{i=1}^{n}(R_i^{'}-\overline{R'})(Q_i^{'}-\overline{Q'})}{\sqrt{\sum_{i=1}^{n}(R_i^{'}-\overline{R'})^2 \sum_{i=1}^{n}(Q_i^{'}-\overline{Q'})^2}} \quad (3)$$

Jika tidak ada rank yang sama, maka digunakan rumus berikut:

$$r_{W2} = 1 - \frac{90\sum_{i=1}^{n}(R_i - Q_i)^2(2n+2-R_i-Q_i)^2}{n(n-1)(n+1)(2n+1)(8n+11)} \quad (4)$$

Jadi telah kita lihat bahwa koefisien korelasi rank yang dicari pada persamaan (2) adalah koefisien korelasi rank yang dibobotі untuk A=1 dan

$$\mathrm{B} = \frac{90\sum_{i=1}^{n}(R_i - Q_i)^2(2n+2-R_i-Q_i)^2}{n(n-1)(n+1)(2n+1)(8n+11)}$$

Dari persamaan (4) jelas bahwa nilai maksimum dari koefisien baru ini diperoleh ketika dua vektor rank sama atau $R_i = Q_i, \forall i = 1,2,\dots,n$, serta nilai minimum koefisien baru diperoleh ketika $Q_i = n + 1 - R_i, \forall i = 1,2,\dots,n$.

3 PROGRAM PCA KLASIK DAN PROGRAM WPCA DENGAN S-PLUS 2000

Berdasarkan algoritma metode PCA klasik dan metode WPCA yang direkomendasikan oleh da Costa dkk (2011), dibuat program untuk kedua metode oleh penulis dengan bantuan software S-PLUS2000. Berikut program untuk kedua metode tersebut:

```
pcaklasik <- function(data, n, p)
{
    dim(data) <- c(n, p)
    m <- matrix(rep(0, p), ncol = p)
    kmc <- matrix(rep(0, n * p), nrow = n, byrow = T)
    sj <- matrix(rep(0, p), ncol = p)
    R <- matrix(rep(0, n * p), nrow = n, byrow = T)
    for(j in 1:p)
    {
        m[, j] <- mean(data[, j])
        kmc[, j] <- (data[, j] - m[, j]) * (data[, j] - m[, j])
        sj[, j] <- sqrt(sum(kmc[, j]/n))
        R[, j] <- (data[, j] - m[, j])/(sj[, j] * sqrt(n))
    }
    c <- t(R) %*% R
    nil <- eigen(c)$values
    v <- eigen(c)$vectors
    list(R, c, nil, v)
}
wpca <- function(data, n, p)
{
    dim(data) <- c(n, p)
    Raksen <- matrix(rep(0, n * p), nrow = n, byrow = T)
    m <- matrix(rep(0, p), ncol = p)
    kmc <- matrix(rep(0, n * p), nrow = n, byrow = T)
    srj <- matrix(rep(0, p), ncol = p)
    Rd <- matrix(rep(0, n * p), nrow = n, byrow = T)
    for(j in 1:p)
    {
        Raksen[, j] <- data[, j] * (2 * n + 2 - data[, j])
        m[, j] <- mean(Raksen[, j])
        kmc[, j] <- (Raksen[, j] - m[, j]) * (Raksen[, j] - m[, j])
        srj[, j] <- sqrt(sum(kmc[, j]/n))
        Rd[, j] <- (Raksen[, j] - m[, j])/(srj[, j] * sqrt(n))
    }
    c <- t(Rd) %*% Rd
    nil <- eigen(c)$values
    v <- eigen(c)$vectors
    list(Raksen, m, kmc, srj, Rd, c, nil, v)
}
```

4 HASIL PENERAPAN PCA KLASIK DAN WPCA

Program untuk metode WPCA yang telah dibuat oleh penulis selanjutnya diterapkan pada data bayi (Damayanti,2008) dan data catatan waktu pelari (Johnson, 2007).

Data pertama yaitu data bayi atau data status gizi ibu hamil dan karakteristik bayi baru lahir (Damayanti, 2008). Data tersebut merupakan data yang diukur pada ibu hamil dan bayinya yang dilahirkan di beberapa kota di daerah JawaTengah. Data tersebut terdiri dari 45 pengamatan dengan 7 variabel yang diukur pada ibu dan bayi yang telah dilahirkannya. Berikut beberapa singkatan dari 7 variabel yang diukur pada ibu dan bayinya : TLI = tebal lemak ibu, BBB = berat badan bayi (gram), PBB = panjang badan bayi (cm), LK = lingkar kepala bayi (cm), LD = lingkar dada bayi (cm), dan LLAI = lingkar lemak atas ibu (mm).

Di Indonesia masalah ibu hamil dan kesehatan bayi baru lahir masih merupakan masalah kesehatan masyarakat yang perlu memperoleh perhatian khusus, tingkat kematian bayi relatif tinggi. Lingkar lemak atas dan tebal lemak merupakan salah satu cara pengukuran terhadap ibu hamil. Ibu hamil yang menderita status gizi kurang mempunyai resiko melahirkan bayi dengan kematian saat persalinan, pendarahan, paska persalinan yang sulit karena lemah dan mudah mengalami gangguan kesehatan kesehatan ibu ketika hamil. Oleh karena permasalahan yang ingin dikaji adalah masalah kurang gizi yang terjadi pada ibu hamil, maka rangking yang tertinggi (rank=1) akan diberikan pada data terkecil pada setiap hasil pengukuran.

Dari tabel 2 dapat dilihat proporsi dari masing-masing Komponen dan proporsi kumulatif. Karena jumlah Komponen yang akan digunakan sebagai Komponen utama memiliki proporsi kumulatif minimal 80% dari total variasi data, maka untuk hasil analisis data tersebut diperoleh sebanyak empat buah Komponen sebagai Komponen utama.

Nilai loading merupakan koefisien dari transformasi Komponen-Komponen utama yang memberikan hasil yang tepat mengenai pengaruh variabel-variabel asli dari Komponen utama, dan merupakan dasar yang bermanfaat untuk interpretasi. Nilai koefisien yang besar menerangkan loading yang tinggi, dan ketika nilai koefisien mendekati nol, artinya Komponen tersebut memiliki loading yang rendah.

Nilai loading untuk ketujuh Komponen dapat dilihat pada tabel 3.

Pendeteksian pencilan dapat menggunakan pendekatan jarak Mahalanobis dan MVE. Pendeteksian pencilan dengan jarak Malahanobis didasarkan pada besar jarak Mahalanobis dibandingkan dengan nilai ${\chi^2}_{\text{TABEL}}\,(\alpha;p)$, yaitu pengamatan yang dianggap pencilan merupakan pengamatan dengan jarak Mahalanobis $\geq {\chi^2}_{\text{TABEL}}\,(\alpha;p)$. Nilai jarak Mahalanobis untuk setiap pengamatan Data Bayi kemudian dibandingkan dengan nilai ${\chi^2}_{\text{TABEL}}\,(0{,}05;7) = 14{,}06714$. Pengamatan-pengamatan yang terdeteksi sebagai pencilan dengan cara ini adalah pengamatan ke-5, 6, 24, dan 34.

Metode yang kedua yaitu *Minimum Volume Ellipsoid (MVE)* yang diperkenalkan oleh Rousseeuw et.all (2002). Menurut Rousseeuw, Croux dan Haesbroek (2002), MVE dapat didefinisikan sebagai elipsoida terkecil yang mencakup paling sedikit h elemen pada himpunan X $=\{x_1,x_2,\ldots,x_n\}$ di R^p dengan penaksir lokasi MVE (T) merupakan pusat elipsoida dan penaksir *scatter* MVE berkoresponden menjadi matriks bentukan. Pasangan T dan S (matriks kovarians sampel) meminimalkan determinan S pada kondisi

$$\{i|(X_i-T)^t\}S^{-1}(X_i-T)\leq c^2\}\geq h$$

dengan h = (n+p+1)/2, dan c adalah konstanta(umumnya dipilih$c=\sqrt{{\chi^2}_{p,alpha}}$), lihat Rousseeuw dan Van Aelst (2009). Dari data bayi diperoleh:

$$\boldsymbol{T(X)} = \langle 25{,}46242 \quad 17{,}58030 \quad 3010{,}60606 \quad 47{,}65152 \quad 33{,}13636 \quad 32{,}65152 \quad 10{,}72727\rangle$$

sedangkan estimasi matriks kovarians sampel menggunakan metode MVE adalah

$$\boldsymbol{S(X)} = \begin{bmatrix} 7{,}1422189 & 8{,}9647055 & 292{,}2704 & 1{,}1774337 & 0{,}9712216 & 0{,}5286837 & 0{,}7019318 \\ 8{,}9647055 & 20{,}794053 & 494{,}7467 & 2{,}5538589 & 0{,}7543324 & 0{,}5812027 & 1{,}5858665 \\ 292{,}27036 & 494{,}74668 & 151680{,}87 & 432{,}7178 & 200{,}85227 & 380{,}37405 & 250{,}6392 \\ 1{,}1774337 & 2{,}5538589 & 432{,}7178 & 2{,}3200758 & 0{,}3927557 & 1{,}0778883 & 0{,}5894886 \\ 0{,}9712216 & 0{,}7543324 & 200{,}8523 & 0{,}3927557 & 1{,}6136364 & 0{,}6740057 & 0{,}3664773 \\ 0{,}5286837 & 0{,}5812027 & 380{,}3741 & 1{,}0778883 & 0{,}6740057 & 1{,}7107008 & 0{,}7926136 \\ 0{,}7019318 & 1{,}5858665 & 250{,}6392 & 0{,}5894886 & 0{,}3664773 & 0{,}7926136 & 0{,}6264205 \end{bmatrix}.$$

Selanjutnya T dan S ini diterapkan pada metode MVE dengan bantuan software S-PLUS2000, dengan perintah cov.mve(databayi), dan diperoleh hasil : pengamatan yang terdeteksi sebagai pencilan dengan metode MVE ini yaitu pengamatan ke-1, 2, 5, 6, 11, 13, 22, 23, 24, 30, 34, 37, dan 43. Oleh karena pencilan dengan metode MVE lebih banyak jumlahnya (13) dari pada jarak Malahanobis, maka hasil yang akan digunakan adalah hasil metode MVE. Selanjutnya perhatikan tabel 4 – 9.

Dengan mengamati tabel 4 sampai dengan tabel 9, dapat dilihat bahwa nilai loading WPCA dengan pencilan (tabel 5) mirip dengan PCA klasik tanpa pencilan (tabel 3), terutama pada Komponen pertama dan kedua. Jumlah Komponen yang diperoleh PCA klasik dan WPCA dengan pencilan sama, namun nilai proporsi kumulatif pada WPCA dengan pencilan (tabel 4) lebih besar daripada proporsi pada PCA klasik dengan pencilan (tabel 2). Sehingga dapat disimpulkan analisis Komponen utama yang dilakukan menggunakan *Weighted* PCA memberikan hasil yang lebih baik (lebih tangguh/*robust*) dibandingkan dengan PCA Klasik.

Perlu dicatat di sini bahwa subruang berdimensi 4 diperoleh dari hasil metode PCA Klasik (dengan pencilan) dan metode WPCA (dengan pencilan). Sedangkan jika tanpa menyertakan pencilan subruang berdimensi 3 direkomendasikan dari hasil metode PCA klasik dan metode WPCA.

Hasil yang sama ditunjukkan pada data catatan waktu pelari (pria), data dapat dilihat pada Johnson (2007:477). Data tersebut terdiri dari 55 observasi (negara) dengan delapan variabel yang terdiri dari X_1 : 100m (detik), X_1 : 200m (detik), X_1 : 400m (detik), X_1 : 800m (menit), X_1 : 1500m (menit), X_1 : 5000m (menit), X_1 : 10000m (menit), dan X_1 : marathon (menit).

Untuk data tersebut, metode MVE mendeteksi bahwa pengamatan ke 5, 9, 12, 16, 22, 23, 26, 28, 33, 34, 35, 36, 41, 42, 46, 51 dan 55 adalah pencilan. Di sini catatan waktu pelari dengan waktu terendah diberi rangking tertinggi (rank=1). Selanjutnya hasil pengolahan data dapat dilihat pada tabel 10 dan 11.

Pada tabel 10 dapat dilihat bahwa subruang berdimensi 2 dapat digunakan untuk menginterpretasikan data catatan waktu pelari (pria) karena persentase proporsi kumulatif (variasi total yang diterangkan oleh dua Komponen utama pertama) lebih dari 80%. Sehingga untuk keperluan pengelompokan variabel (interpretasi data catatan waktu pelari) dapat dilakukan pada subruang berdimensi dua. Hasil pengelompokan yang diperoleh ini nantinya akan menjadi analisa pendahuluan untuk keperluan pengelompokan dalam analisis klaster, pada artikel ini pengelompokan yang terjadi tidak akan dibahas oleh penulis. Hasil yang diperoleh dari metode WPCA dengan pencilan untuk data tersebut:

Komponen utama pertama:

Komp1 = $0.284115\ X_1 + 0.34233\ X_2 + 0.379923X_3 + 0.376596X_4 + 0.371006X_5 + 0.375099X_6 + 0.367569X_7 + 0.320112X_{8.}$

Komponen utama kedua :

Komp2 = $-0.62196\ X_1 - 0.47482X_2 - 0.1263X_3 - 0.04333X_4 + 0.112684X_5 + 0.289117X_6 + 0.352828X_7 + 0.386149X_8$.

Perhatikan kembali tabel 10, hasil metode PCA klasik tanpa pencilan mirip dengan hasil yang diperoleh dari metode WPCA dengan pencilan. Nilai loading (tabel 11) yang diperoleh dari kedua metode (PCA klasik dan WPCA) untuk dua Komponen utama pertama yang diperoleh nilainya hampir sama, dengan menyertakan pencilan maupun tanpa menyertakan pencilan.

5 SIMPULAN

Hal menarik yang diperoleh dari dua contoh data studi kasus terjadi pada data catatan waktu pelari (pria), hasil metode PCA klasik tanpa pencilan mirip dengan hasil yang diperoleh dari metode WPCA dengan pencilan, sedangkan nilai loading yang diperoleh dari kedua metode (PCA klasik dan WPCA) untuk dua Komponen utama pertama nilainya hampir sama.

Demikian pula yang terjadi pada data bayi, nilai loading WPCA dengan pencilan mirip dengan PCA klasik tanpa pencilan, terutama pada Komponen pertama dan kedua. Jumlah Komponen yang diperoleh PCA klasik dengan pencilan dan WPCA dengan pencilan sama, namun nilai proporsi kumulatif pada WPCA dengan pencilan lebih besar daripada proporsi pada PCA klasik dengan pencilan.

Sehingga dapat disimpulkan dari hasil analisa pada kedua data studi kasus, analisis Komponen utama yang diboboti atau *Weighted* PCA (WPCA) memberikan hasil yang lebih tangguh (*robust*) terhadap pencilan dibandingkan dengan PCA Klasik.

REFERENSI

Da Costa, J. F. P., Alonso, H. and Roque, L. (2011). "A Weighted Principal Component Analysis and Its Application to Gene Expression Data". IEEE/ACM Transaction on Computational Biology and Bioinformatics. Vol. 8 No. 1. January 2011.

Da Costa, J. F. P. and Roque, L. (2006). "Limit Distribution for the Weighted Rank Correlation Coefficient, rw2". REVSTAT-Statistical J.,. Vol. 4, No.3, pp. 189-200.

Damayanti, E. (2008). Hubungan Status Gizi Ibu Hamil dengan Berat Badan dan Panjang Badan Bayi Baru Lahir di Puskesmas Mergangsan Yogyakarta. Skripsi. Yogyakarta : Fakultas Kedokteran UGM Yogyakarta.

Hubert,M. and Sangelen, (2004). Robust PCA and Classification in Bioscience. Bioinformatic. Vol. 20 No. 11 2004, Pages 1728–1736.

Johnson and Wichern. (2007). Applied Mutivariate Statistical Analysis. Sixth Edition. New Jersey : Pearson Education, Inc.

Joliffe, I.T. (1986). Principal Component Analysis. New York : Springer-Verlag.

Rousseeuw, PJ. and Van Aelst, S. (2009). Minimum Volume Elipsoid. WIREsComputational Statistics 1:71-82.

S-Plus2000 Professional Release 1. (1998). MathSoft Inc.

LAMPIRAN

Tabel 1. Data Bayi

No	Nama Ibu	LLAI	TLI	BBB	PBB	LKB	LDB	LLAB
1	Ny Rani Almakomala	23.6	12	2100	48	33	28	8
2	Ny Suprihatin	28.6	12.5	3400	49	33	34	12
3	Ny Mudriyati Lestari	20.6	8.5	2650	46	31	31	9.5
4	Ny Susanti	27.2	21.5	3100	49	33	32	11
5	Ny Ika Rosita	36.85	19.55	3600	51	35	34	12
6	Ny Widarsih	28	15	3200	49	53	32	10
7	Ny Yuniarti	26.6	14.6	3000	47	31	31	10

8	Ny Widyastuti	28	18	2900	48	33	31	10
9	Ny Haryanti	26.85	13.75	2900	47	34	33	10
10	Ny Kiky	24	20.25	2200	46	30	29	9.5
11	Ny Nurkhotimah	28	21.8	3100	49	31	33	13
12	Ny Purwati	23.6	16.5	3000	48	33	32.5	11
13	Ny Erni	29.15	26.65	3300	50	35	35	10
14	Ny Sri Parmiyati	32.15	30.4	3750	50	35	33	12
15	Ny Ari Ismiyati	21.25	9.25	3000	47	34	33	10
16	Ny Dwi Yanuar	23.5	18.75	3500	48	33	33.5	12
17	Ny Lusia Febrian	23.5	10.1	3450	48	35	34	11
18	Ny Juminem	27.76	15.1	3200	49	34	33	11
19	Ny Sumini	27.5	16.5	3300	49	33	35	11
20	Ny Dita Nur	25	19	3700	49	34	34	12
21	Ny Margiyah	26.3	15	3050	48	33.5	33	11
22	Ny Salviana	24.2	18.5	3500	49	30	35	12
23	Ny Tri wahyuningsih	24.5	15.5	3150	50	32	31	11.5
24	Ny Pepi Windarti	26	21	3250	51	33	22	12
25	Ny Fitri	26.5	17.5	2350	43	33	30	10
26	Ny Wiji Hatuti	25.25	16.1	3300	47	34	34	12
27	Ny Yeni Rahmawati	26.3	21	2600	46	33	32	10
28	Ny Musiyem	26.2	20	2900	48	34	33	11
29	Ny Tri wahyuni	27.8	21.25	3000	49	33	34	11
30	Ny Sartini	23.8	17	3150	49	36	36	10
31	Ny Yuniatun	24.2	16.5	2800	49	33	32	10
32	Ny Eni Triasih	21.1	12	2150	47	31	31	9
33	Ny Listianti Nurhardian	25.4	15.5	3000	48	34	33	11
34	Ny Nurkhasanah	24	16	2950	40	34	33	11
35	Ny Kresna Muryani	26	20.1	3400	49	33	33.5	11
36	Ny Wuri Suyani	24	19	2950	48	34	33	11
37	Ny Agustin Sri Handayani	25.8	14.5	3500	51	35	34	12
38	Ny Dwi Puspita	23.8	15	2900	45	34	33	11
39	Ny Ari Krisyanti	23	13	2400	45	34	32	10
40	Ny Retno Handayani	25	19.5	3300	50	34	32	10
41	Ny Suminah Istiana	24.5	20	3300	48	34	33	11
42	Ny Eni Indarsih	32.8	26	2900	47	33	33	11
43	Ny Ilah Jazilah	26	23	2900	46	33	32	9
44	Ny Retno Palupi	27.4	24	2900	48	34	33	11
45	Ny Dalilah	23	18	3000	48	32	33	11

Tabel 2. Nilai Eigen dan Proporsi Masing-masing Komponen Utama

Komponen	Nilai Eigen	% Proporsi	% Proporsi Kumulatif
1	2,8614730	40,87819	40,88
2	1,1820942	16,88706	57,765
3	1,0625125	15,17875	72,944
4	0,8595305	12,27901	85,223
5	0,5194205	7,420293	92,643
6	0,3625741	5,17963	97,823
7	0,1523952	2,177074	100

Tabel 3. Nilai Loading Komponen pada PCA Klasik (dengan Pencilan)

Komponen	1	2	3	4	5	6	7
1	0,4152354	0,3929023	-0,264463856	0,259221584	0,01649385	-0,724347196	-0,10544893
2	0,3330842	0,6112704	-0,003747243	0,330696637	-0,05319201	0,631860925	0,06286064
3	0,5185075	-0,2859252	0,096765749	-0,148276099	0,11858135	0,166948301	-0,75902178
4	0,3988282	0,0333811	-0,088121120	-0,620383608	-0,60942394	0,002943783	0,27527147
5	0,1336937	-0,2906181	-0,860348513	-0,002121394	0,28145936	0,212247999	0,18219342
6	0,2604026	-0,5395321	0,124279992	0,632507184	-0,42791011	-0,002143874	0,20608999
7	0,4505962	-0,1164259	0,396573834	-0,128500751	0,59085790	-0,056024103	0,50731951

Tabel 4. Nilai Eigen dan Proporsi Komponen pada WPCA (dengan Pencilan)

Komponen	1	2	3	4	5	6	7
Nilai Eigen	3,064016	1,316146	0,919309	0,719365	0,417429	0,37615	0,187578
% Proporsi Kumulatif	43,77	62,57	75,71	85,98	91,95	97,32	100

Tabel 5. Nilai Loading pada WPCA (dengan Pencilan)

	Komp.1	Komp.2	Komp.3	Komp.4	Komp.5	Komp.6	Komp.7
Komp.1	0,3117225	-0,49456731	0,45081610	0,153124056	-0,58943771	0,28973736	0,012868620
Komp.2	0,2656220	-0,61088960	0,12992438	-0,427909393	0,57124056	0,13776683	-0,104769077
Komp.3	0,4993617	0,15322612	-0,23039536	0,234171273	0,02759930	0,05002789	-0,784841959
Komp.4	0,4250718	-0,12636415	-0,22399524	0,639590360	0,24268368	0,19095620	0,498735105
Komp.5	0,2524650	0,43742381	0,70826466	-0,009243871	0,06570925	0,48870989	0,006517222
Komp.6	0,4177160	0,38523316	0,05790454	-0,283558987	0,20985730	0,68196455	0,290231675
Komp.7	0,4049699	0,05501595	-0,41429078	-0,499042495	-0,46715364	0,39267352	0,199668062

Tabel 6. Nilai Eigen dan Proporsi Komponen pada PCA Klasik (tanpa Pencilan)

Komponen	1	2	3	4	5	6	7
Nilai Eigen	3,845346	1,449729	0,729653	0,420368	0,276445	0,21099	0,06746
% Proporsi Kumulatif	54,93	75,64	86,07	92,07	96,02	99,04	100

Tabel 7. Nilai Loading pada PCA Klasik (tanpa Pencilan)

	Komp.1	Komp.2	Komp.3	Komp.4	Komp.5	Komp.6	Komp.7
Komp.1	0,2818190	-0,59804486941	-0,219791022	-0,26251671	-0,57241680	-0,30471273	-0,1587024
Komp.2	0,2704296	-0,65048830024	-0,004044129	0,19339237	0,40852115	0,40273428	0,3704476
Komp.3	0,4567457	0,18413528779	0,230157768	0,08729450	0,09434366	-0,63521020	0,5333778
Komp.4	0,3701904	-0,00005233426	0,725416762	-0,41070296	0,08869888	0,18992361	-0,3522985
Komp.5	0,3605285	0,23023643468	-0,587801437	-0,54153176	0,41359100	0,03528642	-0,0770750
Komp.6	0,4120242	0,35752250296	-0,089123035	0,09599392	-0,55115882	0,54420838	0,2920910
Komp.7	0,4494925	0,06686061990	-0,137913124	0,64413946	0,11767626	-0,08935541	-0,5811354

Tabel 8. Nilai Eigen dan Proporsi Komponen pada WPCA (tanpa Pencilan)

Komponen	1	2	3	4	5	6	7
Nilai Eigen	3,56708	1,409866	0,76333	0,592175	0,3219488	0,203366	0,14223
% Proporsi Kumulatif	50,96	71,10	82,00	90,46	95,06	97,97	100,00

Tabel 9. Nilai Loading pada WPCA (tanpa Pencilan)

	Komp.1	Komp.2	Komp.3	Komp.4	Komp.5	Komp.6	Komp.7
Komp.1	0,2554248	-0,55726015	-0,39648868	-0,4805494	0,4664793	-0,0201251	-0,13446324
Komp.2	0,2540974	-0,64573861	-0,02872887	0,4238017	-0,3803776	-0,1097762	0,42577581
Komp.3	0,4488425	0,18071213	0,35936498	-0,2379984	0,2157812	0,4492405	0,57595012
Komp.4	0,4142948	-0,11575403	0,51054494	-0,3516266	-0,4346129	-0,1550784	-0,46661097
Komp.5	0,3283424	0,33914205	-0,66860847	-0,1861306	-0,5156817	0,1670048	0,04091278
Komp.6	0,4383092	0,33170044	-0,05032019	0,1483700	0,2733734	-0,7612054	0,13837410
Komp.7	0,4440376	0,03720749	-0,05101616	0,5937353	0,2481923	0,3928785	-0,47996773

Tabel 10. Nilai Eigen dan Persentase Proporsi Kumulatif (Data Catatan Waktu Pelari)

PCA KLASIK DENGAN PENCILAN

	1	2	3	4	5	6	7	8
nilai eigen	6.6148002	0.878501	0.164801	0.124546	0.080006	0.068335	0.04642	0.02259
%propkum	82.685002	93.66626	95.72628	97.28311	98.28319	99.13737	99.71762	100

WPCA DENGAN PENCILAN

	1	2	3	4	5	6	7	8
nilai eigen	5.8137492	1.140036	0.361086	0.292664	0.189153	0.108157	0.063948	0.031208
%propkum	72.671865	86.92231	91.43589	95.09419	97.45859	98.81055	99.6099	100

PCA KLASIK TANPA PENCILAN

	1	2	3	4	5	6	7	8
nilai eigen	5.7823238	1.200382	0.374038	0.301241	0.13148	0.119119	0.070631	0.020785
%propkum	72.279048	87.28382	91.9593	95.72482	97.36832	98.85731	99.74019	100

WPCA TANPA PENCILAN

	1	2	3	4	5	6	7	8
nilai eigen	5.0766008	1.291744	0.615476	0.422125	0.324029	0.144388	0.079807	0.04583
%propkum	63.457509	79.60431	87.29777	92.57433	96.62469	98.42953	99.42713	100

Tabel 11. Nilai Loading untuk Dua Komponen Utama dari Data Catatan Waktu Pelari (Pria)

PCA KLASIK DENGAN PENCILAN			WPCA DENGAN PENCILAN			PCA KLASIK TANPA PENCILAN			WPCA TANPA PENCILAN		
	komp1	komp2		komp1	komp2		komp1	komp2		komp1	komp2
1	0.317632	-0.56556	1	0.284115	-0.62196	1	0.317206	-0.51102	1	0.326649	-0.50677
2	0.337158	-0.46084	2	0.34233	-0.47482	2	0.344714	-0.42378	2	0.356152	-0.44139
3	0.354442	-0.25144	3	0.379923	-0.1263	3	0.377679	-0.19432	3	0.397805	-0.15365
4	0.369006	-0.01304	4	0.376596	-0.04333	4	0.345573	-0.22569	4	0.383847	-0.122
5	0.373015	0.13986	5	0.371006	0.112684	5	0.379469	0.137375	5	0.350043	0.077599
6	0.364488	0.312498	6	0.375099	0.289117	6	0.37497	0.336577	6	0.372742	0.345942
7	0.36696	0.306932	7	0.367569	0.352828	7	0.355342	0.416047	7	0.354404	0.440727
8	0.342035	0.439266	8	0.320112	0.386149	8	0.328061	0.406601	8	0.271661	0.435797