



Penerapan *Principal Component Analysis Biplot* Untuk Memetakan Provinsi di Indonesia Berdasarkan Sarana Pelayanan Kesehatan

Titi Purwandari*, Yuyun Hidayat

Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Padjadjaran, Bandung
*E-mail: titipurwandari@yahoo.com

Abstrak

Kesehatan lingkungan merupakan suatu kondisi lingkungan yang mampu menopang keseimbangan ekologi yang dinamis antara manusia dengan lingkungan dan untuk mendukung tercapainya kualitas hidup manusia yang sehat. Program Lingkungan Sehat bertujuan untuk mewujudkan mutu lingkungan hidup yang sehat melalui pengembangan sistem kesehatan wilayah. Penelitian ini bertujuan untuk memetakan provinsi provinsi di Indonesia berdasarkan sarana pelayanan kesehatan, hal ini dapat memberikan informasi dan rekomendasi kepada pemerintah Indonesia dalam membuat kebijakan kebijakan di sektor kesehatan. Data yang digunakan bersumber dari Direktorat Jenderal Pengendalian Penyakit dan Penyehatan Lingkungan Kementerian Kesehatan Republik Indonesia. Hasil analisis menggunakan *Principal Component Analysis Biplot*, diperoleh peta pengelompokan provinsi provinsi di Indonesia berdasarkan variabel sarana pelayanan kesehatan dan peta yang dihasilkan dapat mewakili informasi pengelompokan 34 provinsi di Indonesia berdasarkan 9 variabel sarana kesehatan sebesar 85,576%.

Kata kunci: Kesehatan Lingkungan, Peta Pengelompokan, Principal Component Analysis Biplot.

1. Pendahuluan

Menurut Himpunan Ahli Kesehatan Lingkungan Indonesia (HAKLI), kesehatan lingkungan merupakan suatu kondisi lingkungan yang mampu menopang keseimbangan ekologi yang dinamis antara manusia dengan lingkungan dan untuk mendukung tercapainya kualitas hidup manusia yang sehat dan bahagia. Program Lingkungan Sehat bertujuan untuk mewujudkan mutu lingkungan hidup yang sehat melalui pengembangan sistem kesehatan wilayah [3]. Menurut WHO (World Health Organization), kesehatan lingkungan ialah suatu keseimbangan ekologi yang harus tercipta diantara manusia dengan lingkungannya agar bisa menjamin keadaan sehat dari manusia. Indikator sehat menurut WHO berkaitan dengan status kesehatan masyarakat yaitu Indikator komprehensif (angka kematian kasar menurun, rasio angka mortalitas proporsional rendah, umur harapan hidup meningkat), indikator spesifik (angka kematian ibu dan anak menurun, angka kematian karena penyakit menular menurun, angka kelahiran menurun), berhubungan dengan pelayanan kesehatan (rasio antara pelayanan kesehatan dan jumlah penduduk seimbang, distribusi tenaga kesehatan merata, informasi lengkap tentang fasilitas kesehatan, informasi tentang sarana pelayanan kesehatan di Rumah Sakit, Puskesmas, dan lain-lain) [3]. Penelitian ini

berfokus pada sarana pelayanan kesehatan. Tujuan penelitian ini adalah memetakan provinsi provinsi di Indonesia berdasarkan sarana pelayanan kesehatan, hal ini dapat memberikan informasi dan rekomendasi kepada pemerintah Indonesia dalam membuat kebijakan kebijakan di sektor kesehatan lingkungan khususnya sarana pelayanan kesehatan..

2. Metode

Metode yang digunakan dalam penelitian ini adalah *Principal Component Analysis Biplots*. Pada bagian ini akan dibahas mengenai data yang digunakan dalam penelitian serta langkah-langkah dalam mengidentifikasi provinsi provinsi di Indonesia berdasarkan sarana pelayanan kesehatan.

2.1 Metode Pengumpulan Data

Data yang dianalisis merupakan data sekunder tahun 2014 yang dikumpulkan oleh Badan Pusat Statistik dengan obyek penelitian adalah seluruh provinsi di Indonesia yang berjumlah 34 provinsi dan variabel penelitian yang terkait dengan sarana pelayanan kesehatan sejumlah 9 variabel yaitu jumlah penduduk, jumlah tenaga medis, jumlah perawat dan bidan, jumlah tenaga farmasi, jumlah tenaga gizi, jumlah dokter gigi, jumlah posyandu, jumlah puskesmas, dan jumlah rumah sakit di setiap provinsi [2].



2.2 Metode Analisis Data

2.2.1 Principal Component Analysis Biplot

Principal Component Analysis Biplot merupakan metode pemetaan dalam analisis multivariat yang memuat informasi dalam sebuah tabel data, yang menunjukkan struktur utama data [1]. Analisis ini bertujuan untuk menyajikan data dalam peta dua dimensi sehingga perilaku data mudah dilihat dan diinterpretasikan. Analisis biplot memerlukan data dari sejumlah objek dengan variabel berskala interval atau rasio. Metode ini didasarkan pada *Singular Value Decomposition (SVD)* dari suatu matriks data yang telah terkoreksi oleh rata-rata. Langkah langkah analisis komponen utama biplot adalah [4]:

2.2.1 Membentuk Matriks Data

Data berupa kabupaten-kota dengan variabel penelitian disajikan ke dalam matriks awal Y yang berukuran $n \times p$.

$$Y = \begin{bmatrix} y_{11} & y_{12} \dots & y_{1p} \\ y_{21} & y_{22} \dots & y_p \\ \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} \dots & y_{np} \end{bmatrix} \quad (1)$$

Matriks Y pada persamaan (1) ditransformasi terhadap rata-rata.

$$X = \begin{bmatrix} x_{11} & x_{12} \dots & x_{1p} \\ x_{21} & x_{22} \dots & x_p \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} \dots & x_{np} \end{bmatrix} \quad (2)$$

2.2.2 Menghitung Eigenvalue dan Eigenvector

Sebelum mencari nilai dekomposisi singularnya (SVD) perlu dihitung *eigenvalue* dan *eigenvector* dari matriks data $X^T X$ [3]. *Eigenvalue* dinotasikan dengan λ dan *eigenvector* yang dinotasikan dengan a dapat dihitung sebagai berikut :

$$|X^T X - \lambda_i I| = 0 \quad (3)$$

$$X^T X - \lambda_i a = 0 \quad (4)$$

2.2.3 Singular Value Decomposition (SVD)

Pendekatan langsung untuk memperoleh nilai dekomposisi singularnya (SVD) adalah sebagai berikut [4]:

$$X_{(n \times p)} = U_{(n \times r)} L_{(r \times r)} A^T_{(r \times p)} \quad (5)$$

dengan :

- $r \leq n, p$

- U dan A^T adalah matriks dengan kolom ortonormal sehingga $U^T U = A^T A = I_r$ (I_r adalah matriks identitas berdimensi r)
- L adalah matriks berukuran $(r \times r)$ dengan unsur-unsur diagonalnya adalah akar kuadrat dari *eigenvalue* $X^T X$, dengan $\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \dots \geq \sqrt{\lambda_r}$ yang membentuk matriks sebagai berikut :

$$L = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\lambda_r} \end{bmatrix} \quad (6)$$

Unsur-unsur diagonal matriks disebut dengan nilai singular matriks X .

- Kolom-kolom matriks A adalah *eigenvector* dari matriks $X^T X$ yang berpadanan dengan *eigenvalue* λ_i yaitu :

$$A = a_1, a_2, \dots, a_r \quad (7)$$

- Kolom-kolom matriks U didapatkan dari rumus :

$$u_i = \frac{1}{\sqrt{\lambda_i}} X a_i, \quad i = 1, 2, \dots, r \quad (8)$$

dengan:

- u_i : unsur-unsur matriks U
- a_i : unsur-unsur matriks A
- λ_i : *eigenvalue* ke- i dari matriks $X^T X$
- X : matriks asal yang dikoreksi terhadap rata-rata

Setelah hasil SVD didapatkan, maka Persamaan 5 dijabarkan menjadi persamaan :

$$X = U L^\alpha L^{1-\alpha} A^T \quad (9)$$

Dalam menentukan L^α , untuk $0 \leq \alpha \leq 1$, maka matriks diagonalnya mempunyai unsur-unsur diagonalnya $\sqrt{\lambda_1^\alpha} \geq \sqrt{\lambda_2^\alpha} \geq \dots \geq \sqrt{\lambda_r^\alpha}$.

Penentuan $L^{1-\alpha}$ berlaku sama dengan unsur-unsur diagonalnya $\sqrt{\lambda_1^{1-\alpha}} \geq \sqrt{\lambda_2^{1-\alpha}} \geq \dots \geq \sqrt{\lambda_r^{1-\alpha}}$.

Misalkan $G = U L^\alpha$ dan $H^T = L^{1-\alpha} A^T$, persamaan (9) menjadi:

$$GH^T = U L^\alpha L^{1-\alpha} A^T = U L A^T = X \quad (10)$$



Unsur ke- (i,j) dalam matriks \mathbf{X} dapat dituliskan sebagai berikut :

$$\mathbf{x}_{ij} = \mathbf{g}_i \mathbf{h}_j \quad (11)$$

Dengan $\mathbf{g}_i, i = 1, 2, \dots, 34$ dan $\mathbf{h}_j, j = 1, 2, \dots, 9$ masing-masing merupakan baris matriks \mathbf{G} dan kolom matriks \mathbf{H} . Pada \mathbf{g}_i dan \mathbf{h}_j mempunyai r dimensi. Dua kolom pertama dari matriks \mathbf{G} dapat digunakan untuk pemetaan objek, sedangkan dua kolom pertama dari matriks \mathbf{H}^T digunakan untuk pemetaan variabel.

2.2.4 Identifikasi Persentase Keragaman Data

Jika matriks \mathbf{X} mempunyai rank lebih dari dua maka *eigenvalue* yang diambil adalah λ_1 dan λ_2 sehingga besarnya keragaman yang diterangkan adalah sebagai berikut :

$$\tau = \frac{(\lambda_1 + \lambda_2)}{\sum_{i=1}^p \lambda_i} \quad (12)$$

dengan :

- λ_1 : *Eigenvalue* terbesar pertama
- λ_2 : *Eigenvalue* terbesar kedua
- λ_i : *Eigenvalue* ke- i dari $\mathbf{X}^T \mathbf{X}$; $i = 1, 2, \dots, p$

Jika nilai τ semakin mendekati nilai 1 berarti biplot yang diperoleh dari matriks pendekatan dengan rank = 2 akan memberikan penyajian yang semakin baik mengenai informasi-informasi yang terdapat pada data yang sebenarnya. Maka berdasarkan nilai τ tersebut peta pengelompokan yang dihasilkan dapat digunakan dalam pengambilan keputusan. Peta pengelompokan yang dihasilkan dapat memberikan gambaran posisi kedekatan satu objek dengan objek lainnya serta kedekatan variabel terhadap objek.

2.2.5 Identifikasi Hasil Pemetaan

Hasil pemetaan dari Biplot *PCA* adalah sebagai berikut [4] :

1. Kedekatan (similaritas) antar objek. Semakin dekat posisi dua buah titik objek maka semakin mirip, semakin jauh posisi dua buah titik objek maka semakin berbeda .
2. Keragaman variabel.
Variabel digambarkan sebagai garis berarah. Variabel dengan keragaman kecil digambarkan sebagai vektor berukuran pendek sedangkan variabel dengan keragaman besar digambarkan sebagai vektor berukuran panjang.

3. Hubungan atau korelasi antar variabel. Hubungan antar variabel dapat diidentifikasi berdasarkan sudut yang terbentuk dari dua buah vector variabel terhadap sumbu peta. Jika dua buah vektor variabel berhimpit terhadap sumbu peta dengan arah yang sama (mendekati 0° atau 360°), maka memiliki korelasi positif sangat erat. Jika dua buah vektor variabel berhimpit dengan sumbu peta dengan arah berlawanan (mendekati 180°), maka memiliki korelasi negatif yang sangat erat, jika dua buah vektor variabel tegak lurus terhadap sumbu peta (mendekati 90° atau 270°) maka dua buah variabel tersebut tidak berkorelasi.
4. Nilai variabel pada suatu objek.
Objek yang terletak searah dengan arah dari suatu variabel, dikatakan bahwa pada objek tersebut nilainya di atas rata-rata. Sebaliknya, jika objek lain terletak berlawanan dengan arah dari variabel tersebut, maka objek tersebut memiliki nilai dibawah rata-rata. Sedangkan objek yang hampir di tengah-tengah, memiliki nilai dekat dengan rata-rata.

3. Hasil dan Pembahasan

Hasil pengolahan data penelitian menggunakan *principal component analysis biplot*, adalah sebagai berikut :

Tabel 1. Identifikasi Persentase Keragaman Data

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	6.594	73.267	73.267
2	1.108	12.309	85.576
3	.692	7.689	93.265
4	.338	3.759	97.024
5	.139	1.542	98.566
6	.071	.785	99.351
7	.041	.457	99.808
8	.010	.114	99.922
9	.007	.078	100.000

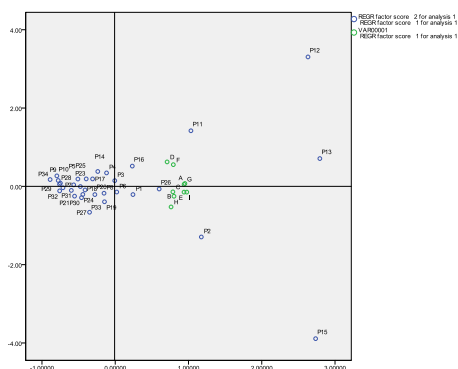
Persentase keragaman data yang dapat dijelaskan oleh peta yang dihasilkan analisis Biplot menggunakan 2 komponen utama (PC1 dan PC2) adalah 85,576%, artinya peta yang dihasilkan dapat mewakili informasi pengelompokan provinsi provinsi



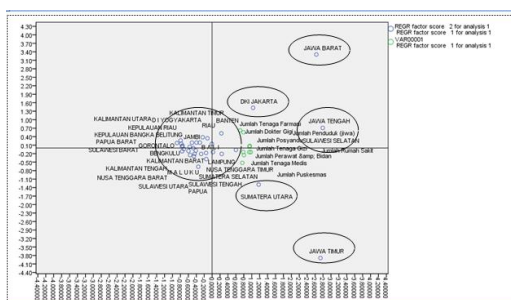
di Indonesia berdasarkan indikator sarana pelayanan kesehatan.

2. Informasi Identifikasi Hasil Pemetaan

Gambar 1 dan Gambar 2 menyajikan informasi 6 kelompok provinsi di Indonesia yang memiliki karakteristik mirip berdasarkan 9 variabel sarana kesehatan, kelompok 1 yaitu provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua, Kalimantan Utara. Kelompok 2 adalah DKI Jakarta, kelompok 3 adalah provinsi Jawa Barat, Kelompok 4 adalah provinsi Jawa Tengah dan Sulawesi Selatan, kelompok 5 adalah provinsi Sumatera Utara, dan kelompok 6 adalah provinsi Jawa Timur.



Gambar 1. Peta Pengelompokan Provinsi di Indonesia Berdasarkan Kode

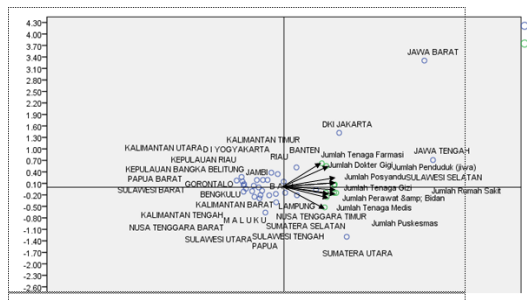


Gambar 2. Peta Pengelompokan Provinsi di Indonesia

3. Berdasarkan Keragaman Variabel

Gambar 3 memperlihatkan bahwa berdasarkan panjang vektor yang dibentuk oleh variabel sarana pelayanan kesehatan, vektor paling panjang ditunjukkan oleh variabel I (Jumlah Rumah Sakit), hal ini menunjukkan bahwa jumlah rumah sakit untuk provinsi provinsi di Indonesia memiliki variasi besar atau beragam, sedangkan

vektor paling pendek ditunjukkan oleh variabel D (jumlah Tenaga Farmasi), hal ini menunjukkan bahwa Jumlah Tenaga Farmasi untuk provinsi provinsi di Indonesia memiliki variasi kecil / keragaman kecil.



Gambar 3. Plot Provinsi di Indonesia Berdasarkan 9 Variabel Sarana Kesehatan

4. Hubungan atau Korelasi Antar Variabel

Gambar 3 memperlihatkan bahwa dari 9 variabel sarana pelayanan kesehatan, variabel yang memiliki vektor searah dan sudut yang dibentuk oleh dua variabel adalah kecil yaitu variabel jumlah penduduk dengan jumlah posyandu, hal ini menunjukkan korelasi sangat erat dan positif artinya terdapat hubungan yang sangat erat antara variabel jumlah penduduk dan jumlah posyandu, semakin banyak jumlah penduduk maka jumlah posyandu semakin banyak. Demikian pula untuk variabel jumlah tenaga gizi dengan jumlah rumah sakit, hal ini menunjukkan korelasi sangat erat dan positif artinya terdapat hubungan yang sangat erat antara jumlah tenaga gizi dengan jumlah rumah sakit, semakin banyak jumlah rumah sakit maka semakin banyak jumlah tenaga gizi.

5. Berdasarkan Kedekatan Provinsi Dengan Variabel Yang Diteliti.

Kelompok 1 yaitu provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua, Kalimantan Utara, berlawanan arah dengan 9 variabel sarana pelayanan kesehatan, hal ini menunjukkan bahwa provinsi provinsi di kelompok 1 dengan jumlah penduduk, jumlah tenaga medis, jumlah perawat dan bidan, jumlah tenaga farmasi, jumlah tenaga gizi, jumlah dokter gigi, jumlah posyandu, jumlah puskesmas, dan jumlah rumah sakit dibawah rata-rata. Sedangkan kelompok 2 adalah DKI Jakarta, kelompok 3 adalah provinsi Jawa Barat, kelompok 4 adalah provinsi Jawa Tengah dan Sulawesi Selatan, kelompok 5 adalah provinsi Sumatera Utara, dan kelompok 6 adalah provinsi



Jawa Timur, searah dengan vektor 9 variabel, hal ini menunjukkan bahwa provinsi provinsi di klaster 2, klaster 3, klaster 4, klaster 5, klaster 6 merupakan klaster klaster yang berjumlah penduduk, jumlah tenaga medis, jumlah perawat dan bidan, jumlah tenaga farmasi, jumlah tenaga gizi, jumlah dokter gigi, jumlah posyandu, jumlah puskesmas, dan jumlah rumah sakit diatas rata rata.

4. Kesimpulan

Berdasarkan hasil analisis data, maka dapat disimpulkan bahwa persentase keragaman data yang dihasilkan oleh *Principal Component Analysis Biplot* adalah sebesar 85,576%, artinya peta yang dihasilkan dapat mewakili informasi pengelompokan 34 provinsi di Indonesia berdasarkan 9 variabel sarana kesehatan sebesar 85,576%.

Kelompok 1 yaitu provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua, Kalimantan Utara, berlawanan arah dengan 9 variabel sarana

pelayanan kesehatan, hal ini menunjukkan bahwa provinsi provinsi di kelompok 1 memiliki jumlah penduduk, jumlah tenaga medis, jumlah perawat dan bidan, jumlah tenaga farmasi, jumlah tenaga gizi, jumlah dokter gigi, jumlah posyandu, jumlah puskesmas, dan jumlah rumah sakit dibawah rata rata. Sedangkan kelompok 2 adalah DKI Jakarta, kelompok 3 adalah provisi Jawa Barat, kelompok 4 adalah provinsi Jawa Tengah dan Sulawesi Selatan, kelompok 5 adalah provinsi Sumatera Utara, dan kelompok 6 adalah provinsi Jawa Timur, searah dengan vektor 9 variabel sarana pelayanan kesehatan, hal ini menunjukkan bahwa provinsi provinsi di klaster 2, klaster 3, klaster 4, klaster 5, klaster 6 merupakan klaster klaster dengan jumlah penduduk, jumlah tenaga medis, jumlah perawat dan bidan, jumlah tenaga farmasi, jumlah tenaga gizi, jumlah dokter gigi, jumlah posyandu, jumlah puskesmas, dan jumlah rumah sakit diatas rata rata.

Daftar Pustaka

- [1] F. Hair, J., A. R., Tatham, R. L., & Black, W. C. *Multivariate Data Analysis*. New Jersey: Prentice Hall International Inc, 2010.
- [2] <https://www.bankdata.depkes.go.id/nasional>
- [3] <https://www.bps.go.id>
- [4] Johnson, Winchern, *Applied Multivariate Statistical Analysis*, Sixth Edition, Prentice Hall International, Inc Upper Saddle River, New Jersey, 2007.