

Penerapan Principal Component Analysis untuk Menentukan Faktor yang mempengaruhi Jumlah Kasus Demam Berdarah Dengue di Sumatera Utara

Riri Syafitri Lubis¹, R. Maisaroh Rezyekiyah Siregar², Mitha Wulandari³,
Silvia Hareni³

^{1,2,3} Universitas Islam Negeri Sumatera Utara, ⁴ STKIP Budidaya
¹riri_syafitri@uinsu.ac.id, ²maisarohsrg@uinsu.ac.id, ³mithawulandary01@gmail.com,
⁴harlenisilvia@gmail.com

Abstract

Dengue Hemorrhagic Fever is a disease caused by a virus that is transmitted by the Aedes mosquito through its bite. From the results of the North Sumatra Provincial Health Office report, the number of dengue cases fluctuated. However, from 2017 to 2019 the number of dengue cases increased. The purpose of this study was to determine the factors that influence the number of dengue cases in North Sumatra based on data in 2020 using Principal Component Analysis. Principal Component Analysis is a multivariate data reduction technique by converting the data into a set of fewer linear combinations, but taking most of the variance from the initial data. Of the 6 independent variables, namely population density, percentage of poverty, number of health facilities, number of health workers, percentage of households with access to proper sanitation, and percentage of slum houses, two new variables were obtained, namely environmental conditions (K1) and housing conditions (K2). Environmental conditions (K1) can explain 3,512 of the total communalities with a percentage of diversity of 58.538%. While the condition of the house (K2) explains 1,663 of the total communalities with a percentage of diversity of 27.718%.

Keyword: Dengue Fever, Principal Component Analysis, Eigenvalues.

Abstrak

Demam Berdarah Dengue adalah penyakit karena virus bawaan nyamuk Aedes melalui gigitannya. Dari hasil laporan Dinas Kesehatan Provinsi Sumut, jumlah kasus DBD berfluktuasi. Namun, dari tahun 2017 hingga 2019 jumlah kasus DBD meningkat. Tujuan penelitian ini adalah mengetahui faktor-faktor yang mempengaruhi jumlah kasus DBD di Sumatera Utara berdasarkan data tahun 2020 menggunakan Principal Component Analysis. Principal Component Analysis merupakan teknik reduksi data multivariat dengan mengubah data menjadi sekumpulan kombinasi-kombinasi linier yang lebih sedikit, Namun mengambil hampir sebagian besar varians dari data awal. Dari 6 variabel bebas yaitu kepadatan penduduk, persentase kemiskinan, jumlah sarana kesehatan, jumlah tenaga kesehatan, persentase rumah tangga dengan akses sanitasi layak, dan persentase rumah kumuh, diperoleh dua variabel baru yaitu kondisi lingkungan (K₁) dan kondisi perumahan (K₂). Kondisi lingkungan (K₁) dapat menjelaskan 3.512 dari total *communalities* dengan persentase keanekaragaman 58,538%. Sedangkan kondisi rumah (K₂) menjelaskan 1.663 dari total communalities dengan persentase keragaman 27,718%.

Kata Kunci: Demam Berdarah, *Principal Component Analysis*, Nilai Eigen.

1. Pendahuluan

Demam Berdarah Dengue (disingkat DBD) merupakan penyakit karena virus bawaan nyamuk bergenus Aedes melalui gigitannya. Virus ini disebut virus Dengue yang merupakan Genus *Flavivirus*, famili *Flaviviridae*. Nyamuk inilah yang berperan sebagai

vektor penyebaran DBD. Virus membutuhkan waktu 4-6 hari masa inkubasi intrinsik di dalam tubuh manusia sebelum menimbulkan penyakit. Demam Berdarah Dengue menimbulkan gejala demam 2-7 hari, Nyeri pada tulang dan otot, sakit kepala, ruam pada kulit yang memerah, dan nyeri dibelakang mata. Demam Berdarah Dengue mengakibatkan pendarahan yang disebabkan kebocoran pembuluh darah, dan dapat mengakibatkan kematian (Mufidah & Purhadi, 2016).

Seiring peningkatan kepadatan dan perpindahan masyarakat, banyaknya penderita DBD juga meningkat dan penyebarannya terus bertambah. Dari hasil laporan Dinas Kesehatan Provinsi Sumatera Utara, jumlah kasus DBD mengalami fluktuasi. Namun, dari tahun 2017 sampai dengan 2019 jumlah kasus DBD meningkat. Pada tahun 2019, *Incident Rate* atau angka kesakitan DBD sebesar 52,1 yang dimana ini termasuk kedalam daerah kategori *Incident Rate* tinggi. Dalam hal ini, kasus DBD telah mendominasi hampir diseluruh kabupaten/kota di Sumatera Utara (Dinas Kesehatan Provinsi Sumatera Utara, 2019).

Berdasarkan pernyataan dari Kemenkes (2017), banyak faktor yang menjadi penyebab penularan dan perkembangan penyakit Demam Berdarah Dengue, termasuk bertambahnya jumlah penduduk, penyediaan air bersih kurang baik, sistem pengelolaan limbah, meluasnya perkembangan dan penyebaran nyamuk, sistem kendali nyamuk kurang efektif dan struktur kesehatan penduduk yang kurang baik. Hal ini tentunya juga disebabkan oleh bertambahnya kemiskinan, kurangnya sarana kesehatan dan tenaga kesehatan, minimnya rumah tangga berakses sanitasi layak, dan bertambahnya pemukiman (perumahan) kumuh.

Principal Component Analysis adalah teknik reduksi data multivariat dengan merubah data pertama menjadi kombinasi-kombinasi linier yang lebih sedikit, namun mengambil hampir bagian besar varian data pertama (Jolliffe, 2002). *Principal Component Analysis* ini sangat cocok untuk mengetahui faktor yang paling mempengaruhi suatu permasalahan, dari sekian banyak faktor yang ada, dengan masih menyerap informasi setiap faktor. Oleh sebab itu, penulis tertarik meneliti tentang “Penerapan *Principal Component Analysis* untuk Menentukan Faktor yang Mempengaruhi Jumlah Kasus Demam Berdarah Dengue di Sumatera Utara”.

2. Metode Penelitian

Principal Component Analysis mereduksi banyaknya variabel asal menjadi beberapa variabel terbaru yang menggambarkan dengan baik varian data awal. Semakin kuat korelasi (baik positif atau negative) antara variabel, maka semakin bagus hasil analisisnya (Johnson dan Wichern, 2007).

Langkah pertama *Principal Component Analysis* adalah mengubah data yang berdistribusi normal umum agar berdistribusi normal baku, kemudian mencari *eigen value* matriks korelasi yaitu r solusi $\lambda_1, \lambda_2, \dots, \lambda_3$ dengan persamaan determinan sebagai berikut

$$|Z^t Z - \lambda I| = 0$$

Nilai eigen matriks korelasi merupakan *trace* matriks $Z^t Z$.

Untuk tiap nilai eigen λ_j terdapat vektor eigen γ_j dengan sistem persamaan homogen:

$$(Z^t Z - \lambda_j I)\gamma_j = 0$$

Vektor eigen solusi $\gamma_j = (\gamma_{1j}, \gamma_{2j}, \dots, \gamma_{rj})$ yang diambil dari sekian banyak penyelesaian yang sebanding untuk setiap j adalah solusi ternormalkan hingga $\gamma_j \gamma_j = 1$. Apabila semua λ_j tidak sama, maka setiap pasangan vektor ciri saling orthogonal. Vektor γ_j membentuk Z menjadi suku-suku komponen utama sebagai berikut:

$$K_j = \gamma_{1j} z_1 + \gamma_{2j} z_2 + \dots + \gamma_{rj} z_r$$

Total kuadrat setiap variabel baru K_j dengan unsur K_{ji} dimana $i = 1, 2, \dots, n$ adalah λ_j (Johnson, 2002).

Penelitian dilakukan di Dinas Kesehatan Provinsi Sumatera Utara dengan alamat Jl. Prof. HM. Yamin Sh, No. 41AA, Perintis, Kecamatan Medan Tim., Kota medan, Sumatera Utara. Penelitian ini adalah penelitian kuantitatif dengan data skunder yang diperoleh dari Dinas Kesehatan Provinsi Sumatera Utara dan data publikasi BPS Provinsi Sumatera Utara. Data diolah menggunakan *software* SPSS dan Matlab. Variabel penelitian adalah

- a. Variabel dependen (Y) yaitu jumlah kasus Demam Berdarah Dengue di Sumatera Utara pada tahun 2020
- b. Variabel independen (X) yaitu :
 - X_1 = Kepadatan Penduduk
 - X_2 = Persentase kemiskinan
 - X_3 = Jumlah sarana kesehatan
 - X_4 = Jumlah tenaga kesehatan
 - X_5 = Persentase rumah dengan sanitasi layak

X_6 = Persentase rumah kumuh

Data yang diperoleh kemudian dianalisis menggunakan *Principal Component Analysis* yang meliputi: pembakuan data, penghitungan koefisien korelasi antar variabel, penghitungan Nilai Eigen, penghitungan komponen-komponen yang direduksi, dan penginterpretasian dengan Model Regresi.

3. Hasil Penelitian dan Pembahasan

Data dari Dinas Kesehatan Provinsi Sumatera Utara dan data publikasi BPS Provinsi Sumatera Utara dibakukan dengan terlebih dahulu mencari rata-rata dan standar deviasinya. Kemudian dilakukan Uji KMO dan Uji Bartlett untuk menentukan apakah data yang digunakan memenuhi persyaratan metode *principle component analysis*. Dengan menggunakan SPSS diperoleh hasil uji Uji KMO sebesar 0,653 dan melebihi batas toleransi kecukupan data yaitu 0,5. Kemudian nilai Bartlett 218,987 dengan signifikan 0 yang lebih kecil dari 0,05. Berarti data berubah (variabel) berhubungan satu sama lain dan cukup memadai untuk melakukan metode *principle component analysis*.

Perhitungan *principle component analysis* dilakukan dengan SPSS. Dengan SPSS diperoleh hasil sebagai berikut:

Tabel 1. Nilai Eigen dan Proporsi Kumulatif

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.512	58.538	58.538	3.512	58.538	58.538
2	1.663	27.718	86.256	1.663	27.718	86.256
3	.638	10.640	96.897			
4	.109	1.815	98.711			
5	.051	.844	99.555			
6	.027	.445	100.000			

Eigen values (nilai eigen) memperlihatkan kepentingan relatif setiap faktor dalam memperhitungkan varians variabel yang dianalisis. Susunan nilai eigen diurutkan dari terbesar ke terkecil, dengan ketentuan nilai eigen lebih kecil 1 tidak digunakan dalam perhitungan jumlah komponen terbentuk.

Tabel 2. Matriks Komponen

	Component	
	1	2
X1	.687	.075

X2	.847	.366
X3	.947	.217
X4	.902	.373
X5	.568	-.806
X6	-.538	.830

Matriks Komponen mendistribusikan peubah (variabel) yang terekstrak pada komponen terbentuk.

Tabel 3. Hasil Rotasi Komponen

	Component	
	1	2
X1	.652	.231
X2	.922	.039
X3	.947	.216
X4	.974	.057
X5	.161	.972
X6	-.124	-.981

Component Matriks hasil rotasi komponen menunjukkan distribusi peubah (variable) dengan nyata dan jelas. Dengan merotasi komponen, angka *factor loading* besar menjadi semakin besar.

Setelah dilakukan *principal component analysis* diperoleh 2 komponen utama. Dengan menggunakan SPSS diperoleh 2 variabel baru yaitu skor dari komponen utama yang terbentuk. Variabel-variabel ini akan menjadi variabel bebas baru (*independent*) dan akan dianalisis menggunakan regresi bersama dengan variabel terikat (*Y*). Untuk model persamaan regresi diasumsikan linear yaitu:

$$\hat{Y} = b_0 + b_1K_1 + b_2K_2$$

dengan konstanta b_0 , b_1 , dan b_2 dapat ditaksir sesuai dengan pengamatan data K_1 , K_2 , dan Y yang telah diperoleh. Koefisien b_0 , b_1 , dan b_2 dapat diperoleh dari persamaan berikut:

$$\sum Y = b_0n + b_1 \sum K_1 + b_2 \sum K_2$$

$$\sum YK_1 = b_0 \sum K_1 + b_1 \sum K_1^2 + b_2 \sum K_1K_2$$

$$\sum YK_2 = b_0 \sum K_2 + b_1 \sum K_2K_1 + b_2 \sum K_2^2$$

Dapat disubstitusikan ke dalam nilai - nilai yang sesuai sehingga diperoleh:

$$7584 = 33b_0 - 3E-05 b_2$$

$$7305,079 = 32b_1 + 6E-06b_2$$

$$3405,187 = -3E-05b_1 + 6E-06b_1 + 32b_2$$

Dalam bentuk matriks, persamaan di atas dinyatakan dengan:

$$\begin{bmatrix} \sum Y \\ \sum YK_1 \\ \sum YK_2 \end{bmatrix} = \begin{bmatrix} n & \sum K_1 & \sum K_2 \\ \sum K_1 & \sum K_1^2 & \sum K_1K_2 \\ \sum K_2 & \sum K_1K_2 & \sum K_2^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

$$\begin{bmatrix} 7584 \\ 7305,079 \\ 3405,187 \end{bmatrix} = \begin{bmatrix} 33 & 0 & -3E-05 \\ 0 & 32 & 6E-06 \\ -3E-05 & 6E-06 & 32 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 33 & 0 & -3E-05 \\ 0 & 32 & 6E-06 \\ -3E-05 & 6E-06 & 32 \end{bmatrix}^{-1} \begin{bmatrix} 7584 \\ 7305,079 \\ 3405,187 \end{bmatrix}$$

Untuk mencari nilai b_0 , b_1 , dan b_2 digunakan metode matriks yang dihitung menggunakan MATLAB, dan diperoleh

$$b_0 = 229,818$$

$$b_1 = 228,284$$

$$b_2 = 106,412$$

Jadi, persamaan regresinya adalah:

$$\hat{Y} = 229,818 + 228,284K_1 + 106,412K_2$$

Dengan menggunakan SPSS didapat tabel analisis Varian berikut:

Tabel 4. Analisis Variansi

Model	Sum of Squares	Df	Mean Square	F	Sig.
1 Regression	2029984.258	2	1014992.129	26.780	.000 ^a
Residual	1137021.802	30	37900.727		

Total	3167006.061	32			
-------	-------------	----	--	--	--

Dari tabel 4 diperoleh $F_{hitung} = 26,780$. Dengan $\alpha = 0,01$, jumlah variabel independen 2 dan jumlah data 33, diperoleh $F_{tabel}(2,33) = 5,31$. Karena $F_{hitung} > F_{tabel}$ maka H_0 ditolak, berarti ada pengaruh komponen K_1 dan K_2 terhadap Y .

Dari hasil penelitian dapat dilihat angka *eigenvalues* yang melebihi 1 ada 2, berarti hanya dua komponen yang telah terbentuk. Sedangkan angka *eigenvalues* untuk keempat komponen lainnya di bawah satu. *Eigenvalue* komponen 1 sebesar 3,512 artinya komponen 1 menjelaskan 3,512 dari total *communalities* dengan persentase keragaman 58,538%. *Eigenvalue* komponen 2 sebesar 1,663, artinya komponen 2 menjelaskan 1,663 dari total *communalities* dengan persentase keragaman sebesar 27,718%.

Berdasarkan Tabel 2 dapat dilihat hanya dua komponen terbentuk dari 6 variabel. Ini memperlihatkan bahwa dua komponen adalah jumlah teroptimal dalam mereduksi 6 variabel independen (bebas).

Dalam Tabel 3 diperlihatkan nilai tertinggi variabel (peubah) X_1 , X_2 , X_3 , dan X_4 ada pada komponen 1, maka variabel (peubah) X_1 , X_2 , X_3 , dan X_4 masuk pada komponen 1. Sedangkan variabel X_5 , dan X_6 nilai tertinggi terdapat pada komponen 2, maka variabel X_5 , dan X_6 masuk ke dalam komponen 2.

Kemudian dilakukan penamaan komponen. Penamaan komponen dilakukan dengan cara membuat nama mewakili nama-nama pembentuk komponen tersebut. Berdasarkan hasil rotasi komponen, penamaan komponen yang mempengaruhi jumlah kasus Demam Berdarah Dongue (DBD) di Sumatera Utara Tahun 2020 adalah sebagai berikut:

1. Komponen 1 terbentuk dari 4 peubah (variable), yaitu kepadatan penduduk (X_1), persentase kemiskinan (X_2), jumlah sarana kesehatan (X_3), dan jumlah tenaga kesehatan (X_4),. Sehingga dinamakan kondisi lingkungan
2. Komponen 2 terdiri dari 2 variabel, yaitu persentase rumah tangga dengan akses sanitasi layak (X_5), dan persentase rumah kumuh (X_6), Sehingga dinamakan kondisi rumah.

4. Kesimpulan

Dari penelitian diperoleh kesimpulan yaitu

1. Dengan metode *Principle Component Analysis*, dari 6 variabel independen yaitu kepadatan penduduk (X_1), persentase kemiskinan (X_2), jumlah sarana kesehatan (X_3),

jumlah tenaga kesehatan (X_4), persentase rumah tangga dengan akses sanitasi layak (X_5), dan persentase rumah kumuh (X_6), diperoleh dua variabel baru yaitu kondisi lingkungan (K_1) dan kondisi rumah (K_2). Kondisi lingkungan (K_1) dapat menggambarkan 3,512 dari total *communalities* dengan persentase keragaman sebesar 58,538%. Sedangkan kondisi rumah (K_2) menggambarkan 1,663 dari total *communalities* dengan persentase keragaman sebesar 27,718%.

2. Model regresinya adalah

$$\hat{Y} = 229,818 + 228,284K_1 + 106,412K_2$$

Model regresi memperlihatkan bahwa

- a. Jika kondisi lingkungan (K_1) = 0 dan kondisi rumah (K_2) = 0, maka jumlah kasus Demam Berdarah Dengue (Y) akan bernilai 229,818.
- b. Untuk setiap kondisi lingkungan (K_1) bertambah sebesar satu, akan mengakibatkan meningkatnya maka jumlah kasus Demam Berdarah Dengue (Y) sebesar 228,282.
- c. Untuk setiap kondisi rumah (K_2) bertambah sebesar satu, akan mengakibatkan meningkatnya maka jumlah kasus Demam Berdarah Dengue (Y) sebesar 106,412.

5. Referensi

- Dinas Kesehatan Provinsi Sumatera Utara. (2019). *Profil Kesehatan Provinsi Sumatera Utara*. Sumatera Utara: Dinas Kesehatan Provinsi Sumatera Utara.
- Jolliffe, I. T. (2002). *Principal Component Analysis (Second Edition)*. United States of America: Springer.
- Johnson, R. A. (2002). *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Education International.
- Johnson & Wichern. (2007). *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall.
- Kementrian Kesehatan RI. (2017). *Pedoman Pencegahan dan Pengendalian Demam Berdarah Dengue di Indonesia*. Direktorat Jenderal Pencegahan dan pengendalian Penyakit.
- Mufidah, A. S., & Purhadi. (2016). Analisis Survival Pada pasien Demam Berdarah Dengue DBD di RSUD Haji Surabaya Menggunakan Model Regresi Weibull. *Jurnal Sains dan seni ITS*, 5(2), D301-D306.