

Estimasi, Vol. 2, No.1, Januari, 2021, Hal.11-18
DOI: 10.20956/ejsa.v2i1.11943

Penerapan *Principal Component Analysis* dalam Penentuan Faktor Dominan Cuaca Terhadap Penyebaran Covid-19 di Surabaya

Khusnia Nurul Khikmah¹

¹Departemen Matematika, Fakultas MIPA, Universitas Negeri Surabaya, Surabaya, 60231, Indonesia

* Corresponding author, email: khusnia.17030214054@mhs.unesa.ac.id

Abstract

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by the acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the transmission can mediate human-to human by enviroment. According to Indonesian Meterological, Climatological, and Geophysical Agency found that weather and climate were supporting factors of COVID-19 outbreak so, research and analysis is carried out regarding the most factor were supporting the spread of COVID-19. In this study, using secondary data obtained from data reported by Indonesian Meterological, Climatological, and Geophysical Agency. According the aims of this study by using *Principal Component Analysis* (PCA) there are three principal components which represents the most factor were supporting the spread of COVID-19 they are temperature, humidity, and length of sunshine.

Keywords: Covid-19, Factor, *Principal Component Analysis*, Surabaya, Weather.

Abstrak

Penyakit Coronavirus 2019 (COVID-19) merupakan penyakit menular yang disebabkan oleh sindrom pernapasan akut corona virus 2 (SARS-CoV-2) dan perantara penularannya yaitu antarmanusia dengan melewati lingkungan. Menurut Badan Meteorologi, Klimatologi, dan Geofisika Indonesia menemukan bahwa cuaca dan iklim merupakan faktor pendukung wabah COVID-19 sehingga dilakukan penelitian dan analisis mengenai faktor yang paling mendukung penyebaran COVID-19. Dalam penelitian ini menggunakan data sekunder yang diperoleh dari data yang dilaporkan oleh Badan Meteorologi, Klimatologi, dan Geofisika Indonesia. Berdasarkan tujuan dari penelitian ini dengan menggunakan *Principal Component Analysis* (PCA) terdapat tiga komponen utama yang mewakili faktor paling banyak mendukung penyebaran COVID-19 yaitu suhu, kelembaban, dan lama sinar matahari.

Kata Kunci: Covid-19, Cuaca, Faktor, *Principal Component Analysis*, Surabaya.

1. Pendahuluan

Coronavirus adalah keluarga besar virus yang menyebabkan penyakit pada manusia dan hewan. Pada Desember 2019 ditemukan virus corona jenis baru di Wuhan China, yang kemudian diberi nama *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) yang menyebabkan *Coronavirus Disease-2019* [1]. COVID-19 dapat menyebar melalui tetesan kecil dari hidung atau mulut saat batuk atau bersin [1]. Infeksi COVID-19 juga dapat menyebar jika seseorang menyentuh permukaan yang terinfeksi dan kemudian menyentuh mata, hidung, atau mulut [2].

Estimasi: Journal of Statistics and Its Application
e-ISSN: 2721-3803, p-ISSN: 2721-379X
<http://journal.unhas.ac.id/index.php/ESTIMASI>

Masa inkubasi COVID-19 diperkirakan 14 hari, dengan sebagian besar kasus terjadi sekitar empat hingga lima hari setelah terpapar [3]. Menurut Kementerian Kesehatan RI, tidak ada batasan usia bagi orang yang dapat terinfeksi COVID-19 [1]. *Centers for Disease and Prevention* China melaporkan bahwa sekitar 44.500 kasus COVID-19 telah dikonfirmasi, dengan 87% berusia antara 30 dan 79 tahun [4]. COVID-19 menular dengan perantara manusia ke manusia melalui lingkungan [5]. Indonesia khususnya Surabaya memiliki iklim tropis dan daerah tropis juga cocok untuk penyebaran COVID-19 [6]. Berdasarkan hasil kajian Badan Meterologi, Klimatologi, dan Geofisika (BMKG) Indonesia ditemukan bahwa cuaca dan iklim merupakan faktor pendukung terjadinya wabah COVID-19 [7]. Jadi, masuk akal jika dilakukan penelitian dan analisis mengenai faktor yang paling mendukung penyebaran COVID-19.

Principal Component Analysis (PCA) merupakan salah satu teknik *multivariate* untuk menganalisis data [8]. Tujuan PCA adalah untuk mengekstrak informasi penting dari data dan mengaturnya ke variabel baru yang disebut komponen utama. PCA juga digunakan untuk meringkas data dengan jumlah variabel yang lebih kecil [9]. Dalam penelitian ini akan dilakukan perhitungan untuk mengetahui faktor dominan yang mendukung penyebaran COVID-19.

2. Material dan Metode

2.1 *Principal Component Analysis*

Principal Component Analysis (PCA) merupakan salah satu teknik *multivariate* untuk menganalisis data [8]. Tujuan PCA adalah untuk mengekstrak informasi penting dari data dan mengaturnya ke variabel baru yang disebut *principal component* (komponen utama). PCA juga digunakan untuk meringkas data dengan jumlah variabel yang lebih kecil [9].

Principal Component Analysis mempunyai beberapa fungsi, yaitu:

- a. Mengidentifikasi variabel baru dari data *multivariable*.
- b. Mereduksi jumlah variabel asli menjadi variabel independen tetapi tetap mempertahankan informasi dari data asli.
- c. Mereduksi variabel asli yang memiliki informasi relatif sedikit dari data asli [10].

Misalkan $X = (X_1, X_2, \dots, X_p)$ adalah vektor variabel asli dan ditransformasikan menjadi vektor variabel baru $K = (K_1, K_2, \dots, K_q)$ dimana $q < p$, persamaan tersebut dapat ditulis dengan:

$$K_i = b_{1i}X_1 + b_{2i}X_2 + b_{3i}X_3 + \dots + b_{pi}X_p \quad (1)$$

Dimana $b_{1i}^2 + b_{2i}^2 + b_{3i}^2 + \dots + b_{pi}^2 = 1$.

Principal Component Analysis dari (1) dapat ditulis menjadi persamaan (2):

$$\begin{aligned} K_1 &= b_{11}X_1 + b_{21}X_2 + b_{31}X_3 + \dots + b_{p1}X_p \\ K_2 &= b_{12}X_1 + b_{22}X_2 + b_{32}X_3 + \dots + b_{p2}X_p \\ &\vdots \end{aligned}$$

$$K_q = b_{1q}X_1 + b_{2q}X_2 + b_{3q}X_3 + \dots + b_{pq} \quad (2)$$

Dimana:

p = banyaknya variabel awal

K_i = komponen ke- i , $i = 1, 2, \dots, q$

b_{pq} = pembebanan variabel p ke- q [8], [11].

2.2 Matriks Masukan

Dalam [11] matriks varians digunakan jika semua variabel satuan ukur adalah sama dan matriks korelasi digunakan bila semua variabel satuan ukur berbeda.

a. Matriks varians-kovarians

Pengukuran sebaran data disebut *variance* dan *covariance* adalah untuk mengetahui seberapa kuat hubungan antara dua variabel. Nilai *variance-covariance* dapat diperoleh dengan mengikuti persamaan:

$$S_{jj} = \frac{\sum_{j=1}^p (x_{ij} - \bar{x}_j)^2}{(p-1)} \quad (3)$$

$$S_{ij} = \frac{\sum_{j=1}^p (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_j)}{(p-1)} \quad (4)$$

Matriks varians-kovarian baru dapat ditulis dengan persamaan berikut:

$$S_{p \times p} = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{bmatrix} \quad (5)$$

Dimana:

S_{jj} = nilai varians

S_{ij} = nilai kovarians

S = matriks varian-kovarian [11], [12].

b. Matriks Korelasi

Matriks dengan elemen diagonal utama berjumlah 1 disebut matriks simetris. Korelasi (r) bertujuan untuk mengukur kekuatan hubungan linier antara dua variabel. Nilai korelasi dapat dilambangkan dengan persamaan:

$$r = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} = \frac{\sum_{j=1}^p (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_j)}{\sqrt{\sum_{j=1}^p (x_{ij} - \bar{x}_i)^2 \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2}} \quad (6)$$

Matriks korelasi dapat ditulis dengan mengikuti persamaan berikut:

$$R_{p \times p} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix} \quad (7)$$

Dimana:

r = variabel koefisien korelasi dari X_i dan X_j

R = matriks korelasi [11], [12].

2.3 Koefisien Pembobot Komponen Utama

Penafsiran variabel komposit yang dihasilkan adalah tujuan dari pembebanan *principal component* (komponen utama). Dari persamaan (2.1) untuk menentukan vektor koefisien $b_i = b_{1i}, b_{2i}, b_{3i}, \dots, b_{pi}$ dapat diselesaikan dengan menggunakan fungsi lagrange:

$$L = b_i^T X^T X b_i - \lambda (b_i^T b_i - 1) \quad (8)$$

dengan syarat $b_i^T b_i = 1$ untuk $b_i^T X^T X b_i$ mencapai nilai maksimum.

Jika L diturunkan oleh vektor b dan sama dengan nol, maka:

$$\begin{aligned} \frac{\partial L}{\partial b} &= 2b_i X^T X - 2\lambda b_i = 0 \\ 2(X^T X - \lambda I)b_i &= 0 \\ (X^T X - \lambda I)b_i &= 0 \end{aligned} \quad (9)$$

Hasil dari $(X - \lambda I)b_i = 0$ adalah non-trivial dan hanya ada satu solusi:

$$|X^T X - \lambda I| = 0 \quad (10)$$

Persamaan (10) akan dihasilkan akar karakteristik λ_i dimana $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_q > 0$ dan selanjutnya setiap akar karakteristik λ_i akan ditentukan b_i [11].

2.4 Penentuan *Principal Component*

Untuk menentukan *principal component* (komponen utama) dapat mengikuti kriteria sebagai berikut:

a. Proporsi Kumulatif Varians Total

Proporsi kumulatif dari varians total dapat ditentukan dengan memilih k *principal component* (komponen utama) sebagai kontribusi terbesar dari varians data dan mempunyai proporsi kumulatif lebih dari 75%.

$$\frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^p \lambda_j} > 0.75 \quad (11)$$

Dimana:

$\sum_{j=1}^k \lambda_j$ = varians komponen utama ke- j , $j = 1, 2, \dots, k$

$\sum_{j=1}^p \lambda_j$ = varians total, $j = 1, 2, \dots, p$.

b. Akar Karakteristik

Komponen yang dipilih memiliki karakteristik root lebih besar dari satu [11].

2.5 Metode Penelitian

Analisis dalam penelitian dilakukan dengan cara:

a. Studi Literatur

b. Persiapan data

Data dalam penelitian ini menggunakan sekunder dari situs resmi Badan Meteorologi, Klimatologi, dan Geofisika Indonesia. Data diambil dari tanggal 1 April 2020 sampai dengan 14 September 2020

c. Menentukan variabel penelitian

Variabel dalam penelitian ini adalah:

1. X_1 = suhu
 2. X_2 = kelembapan
 3. X_3 = lama penyinar matahari
 4. X_4 = kecepatan angin
- d. Melakukan *Principal Component Analysis* dengan:
1. Menentukan matriks korelasi;
 2. Menentukan akar karakteristik dan vektor karakteristik;
 3. Tentukan banyaknya *principal component* (komponen utama);
 4. Menentukan nilai komponen pembobot.
- e. *Ploting* PCA3D.

3. Hasil dan Diskusi

3.1 *Principal Component Analysis*

3.1.1 Matriks Korelasi

Korelasi digunakan untuk mengukur kekuatan hubungan linier dua variable atau lebih. Misalkan R adalah matriks korelasi dari faktor cuaca dan iklim yang akan digunakan untuk mendapatkan koefisien pembobot komponen utama maka dengan bantuan *software* Rstudio diperoleh:

$$R = \begin{bmatrix} 1,0000000 & -0,0020629 & 0,112718 & 0,137543 \\ -0,0020629 & 1,0000000 & 0,060553 & -0,460531 \\ 0,1127182 & 0,0605527 & 1,000000 & -0,026185 \\ 0,1375433 & -0,4605311 & -0,026185 & 1,000000 \end{bmatrix}$$

Koefisien pembobot komponen utama bisa didapatkan dengan terlebih dahulu melakukan Uji Bartlett untuk menganalisis bahwa matriks korelasi R bukan matriks identitas dengan H_0 : matriks korelasi = matriks identitas dan H_1 : matriks korelasi bukan matriks identitas. Hasil Uji Bartlett dengan Rstudio yang didapat bahwa p-value yang diperoleh yaitu mendekati 0 yaitu 0,0000000335 sehingga menolak hipotesis null (H_0) artinya matriks korelasinya bukan matriks identitas. Jadi, analisis faktor cuaca dan iklim dengan menggunakan *Principal Component Analysis* dapat dilanjutkan karena matriks korelasinya bukan matriks identitas.

3.1.2 Koefisien Pembobot dan Jumlah *Principal Component*

Nilai koefisien pembobot diperoleh melalui matriks korelasi untuk mengetahui masing-masing kontribusi variabel ke masing-masing komponen. Sebelum menemukan nilai pembobot nilai *eigen value* dari matriks korelasi terlebih dahulu dicari. Misalkan L adalah nilai *eigen* dari matriks korelasi, nilai L adalah:

$$L = 1,48414 \ 1,12302 \ 0,87292 \ 0,51993$$

Sedangkan nilai pembobotnya yaitu:

	Suhu	Kelembapan	lama penyinaran matahari	kecepatan angin ⁴
Suhu	0,2245	0,7353	0,6230	- 1,44e - 01
Kelembapan	- 0,8294	0,1638	0,2153	4,89e - 01
lama penyinaran matahari	- 0,0979	0,7436	- 0,6614	1,22e - 06
kecepatan angin	0,8580	0,0507	- 0,0303	5,10e - 01

Dari pembobot, kita dapat mengetahui *principal component* (komponen utama) yang memberikan dampak besar ke semua variabel yaitu suhu, kelembapan, dan lama penyinaran matahari:

Suhu	Kelembapan	lama penyinaran matahari
0,2245	0,7353	0,6230
-0,8294	0,1638	0,2153
-0,0979	0,7436	-0,6614
0,8580	0,0507	-0,0303

Selain dari melihat nilai yang dihasilkan dari pembobot komponen pemberi dampak terbesar ke semua variabel juga bisa dilihat dari *scree plot* seperti di bawah yaitu dari *component number* yang berada di atas limit ada dua yaitu *component number* 1, 2, dan 3.

3.1.3 Nilai dari Pembobot Komponen

Misalkan l adalah nilai pembobot (*loading*) komponen utama dari hasil Principal component analysis. nilai l adalah:

$$l = 1,48 \ 1,12 \ 0,87 \ 0,52$$

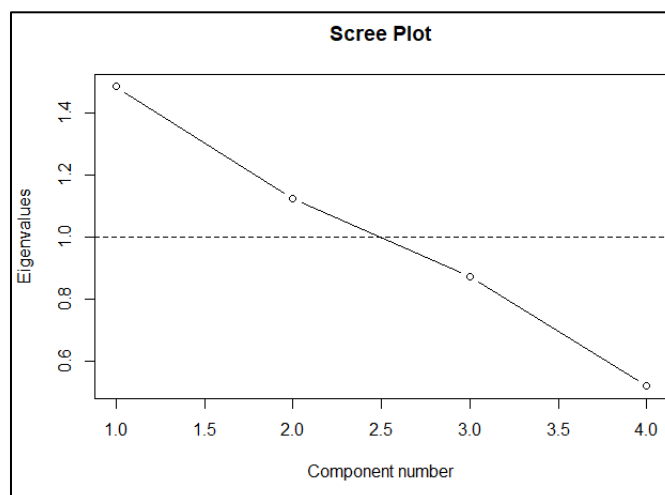
dengan proporsi kumulatif ragam komponen, misalkan CV :

$$CV = 0,37 \ 0,65 \ 0,87 \ 1,00$$

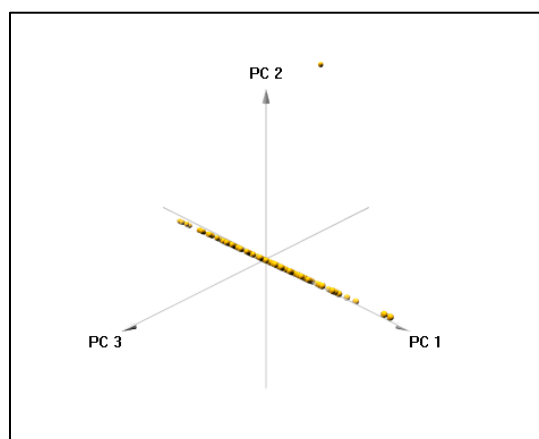
Total kumulatif ragam dari *principal component* (komponen utama) yaitu suhu, kelembapan, dan lama penyinaran matahari adalah sebesar 87%, artinya sebesar 87% dari data terakomodasi pada ketiga komponen tersebut dan memenuhi persamaan (11).

3.2 Grafik PCA3D

Principal Component Analysis dengan tiga faktor dominan menghasilkan grafik PCA3D seperti Gambar 2. Berdasarkan hasil *Principal Component Analysis* diperoleh bahwa suhu, kelembapan, dan lama penyinaran matahari adalah faktor dominan cuaca yang berpengaruh besar terhadap penyebaran COVID-19 di Surabaya oleh karena itu, hasil analisis ini dapat digunakan sebagai bahan pertimbangan dalam membuat keputusan terhadap persebaran COVID-19 berdasarkan kondisi cuaca di Surabaya.



Gambar 1. *Scree plot Principal Component Analysis*



Gambar 2. Grafik pca3d komponen utama

4. Kesimpulan

Berdasarkan *Principal Component Analysis*, kesimpulan dari penelitian ini adalah:

1. Hasil *Principal Component Analysis* dipilih tiga *principal component* (komponen utama) yang memberikan informasi paling banyak (varians) dari data.
2. Tiga komponen yang paling bervariasi pada data adalah suhu, kelembaban dan lama sinar matahari.

Daftar Pustaka

- [1] K. K. R. Indonesia. <https://www.kemkes.go.id/folder/view/full-content/structure-faq.html> (accessed Jul. 26, 2020).
- [2] N. Van Doremalen *et al.* Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.*, 382 (16) : 1564–1567, 2020.
- [3] S. Poudel, Knowledge and Attitudes of Adults in Jhapa District Towards Coronavirus Disease 2019 (COVID-19), 2020.
- [4] Z. Wu & J. M. McGoogan, Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *Jama*, 323 (13) : 1239–1242, 2020.
- [5] M. M. Sajadi, P. Habibzadeh, A. Vintzileos, S. Shokouhi, F. Miralles-Wilhelm & A. Amoroso. “Temperature and latitude analysis to predict potential spread and seasonality for COVID-19,” *Available SSRN 3550308*, 2020.
- [6] M. B. Araujo & B. Naimi. Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate. *medRxiv*, 2020.
- [7] M. Wareza. BMKG:Suhu Dan Kelembapan Tak Ideal Bagi Penyebaran Covid-19. [Online]. Available: <https://www.cnbcindonesia.com/news/20200404143119-4-149777/bmkg-suhu-kelembapan-ri-tak-ideal-bagi-penyebaran-covid-19>, 2020.
- [8] H. Abdi & L. J. Williams. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.*, 2 (4) : 433–459, 2010.

- [9] I. T. Jolliffe. Principal component analysis. *Technometrics*, 45 (3) : 276, 2003.
- [10] F. Solimun, AAR & Nurjannah. *Multivar. Stat. Method Struct. Equ. Model. Based WarpPLS*. 2017.
- [11] S. Astutik. *Analisis Multivariat: Teori dan Aplikasinya dengan SAS*. Universitas Brawijaya Press, 2018.
- [12] R. A. Johnson & D. W. Wichern, *Applied multivariate statistical analysis*, 5 (8). Prentice hall Upper Saddle River, NJ, 2002.