

PENGUNAAN METODE ANALISIS KOMPONEN UTAMA UNTUK MEREDUKSI FAKTOR-FAKTOR DATA

Arafi Ramadhan Maulana (122450002)¹⁾, Dwi Ratna Anggraeni (122450008)²⁾, Raid Muhammad Naufal (122450027)³⁾, Rayan Koemi Karuby (122450038)⁴⁾, Muhammad Deriansyah Okutra (122450101)⁵⁾

Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera

Email :

arafi.122450002@student.itera.ac.id¹, dwi.122450008@student.itera.ac.id²,
raid.122450027@student.itera.ac.id³, rayan.122450038@student.itera.ac.id⁴,
mderiansyah.122450101@student.itera.ac.id⁵

1. Pendahuluan

Penggunaan dari metode analisis komponen utama telah menjadi umum dalam banyak cabang ilmu, mulai dari ilmu sosial, ilmu alam, kedokteran, ekonomi, dan masih banyak cabang ilmu lainnya. Metode analisis komponen utama dapat menjadi jalan keluar yang efektif dalam mencari informasi penting dari data yang rumit. Analisis komponen utama bekerja dengan cara melakukan reduksi dimensi data, yang bertujuan untuk membantu pengguna untuk memahami pola dan hubungan antar variabel. Salah satu contoh sederhana dari pengaplikasian metode analisis komponen utama adalah metode analisis komponen utama membantu membuat portofolio investasi suatu perusahaan dengan menyajikan hubungan dari tiap komponen aset keuangan perusahaan tersebut.

Analisis komponen utama mempunyai beragam pengaplikasian yang luas dan relevan dengan berkembangnya zaman dengan menunjukkan keunggulan dalam menganalisis persoalan statistik dan menyajikan pola yang kompleks dari sebuah data. Dengan dilakukan analisisnya terkait pengaplikasian dari metode analisis komponen utama, diharapkan dengan ditulisnya artikel ini dapat memperluas pandangan pembaca terkait manfaat dan potensi dari metode analisis komponen utama dalam berbagai cabang ilmu.

2. Metode

2.1. Numpy

Numpy merupakan pustaka yang ada pada python yang digunakan untuk memudahkan dalam melakukan perhitungan saintifik seperti matriks, aljabar, statistik, dan sebagainya. Beberapa kelebihan numpy seperti kemampuan komputasi yang berorientasi array, efisien dalam implementasi array multidimensi, terdapat fungsi bawaan untuk aljabar linear dan membuat bilangan acak.

2.1.1. *mean()*

Mean merupakan rata-rata atau jumlah dari keseluruhan data yang ada dan dibagi dengan banyaknya jumlah data.

$$\bar{x} = \frac{\sum x}{n}$$

Keterangan:

\bar{x} = rata-rata

x = nilai data

n = jumlah data

2.1.2. *svd()*

SVD (*Singular value Decomposition*) merupakan pendekatan dekomposisi matriks menjadi tiga komponen yaitu matriks U , matriks Σ , dan matriks V^T . untuk mencari SVD pada python dapat menggunakan fungsi `svd()` dari paket `scipy.linalg`. SVD dapat dirumuskan sebagai berikut jika A merupakan matriks dengan ukuran $m \times n$ sebagai berikut:

$$A = U\Sigma V^T$$

Keterangan:

U = matriks uniter berukuran $m \times m$

Σ = matriks berukuran $m \times n$

V^T = matriks berukuran $n \times n$

2.1.3. *diag()*

Fungsi *diag()* merupakan fungsi dari library Numpy dan biasa digunakan untuk mengekstrak diagonal dari sebuah matriks atau array. Fungsi *diag()* memudahkan pengguna untuk mendapatkan informasi yang diperlukan dari sebuah matriks. Dalam pengaplikasiannya pada kode pemrograman aplikasi kali ini, *diag()* digunakan untuk melihat diagonal matriks dari matriks SVD.

2.2. Pandas

Pandas adalah suatu pustaka dalam Python yang seringkali digunakan untuk melakukan manipulasi pada suatu *dataset*. Pandas dapat membantu dalam memberikan struktur data yang kuat pada *dataset* atau suatu *DataFrame*. Pada kode pemrograman aplikasi kali ini, Pandas digunakan untuk *import dataset* yang akan digunakan dan elemennya dinormalisasi.

2.3. *pca()*

Fungsi *pca()* merupakan fungsi yang ditentukan atau dibuat oleh user (*User-Defined Function*) yang menerima parameter berupa matriks data dengan bentuk (*n_samples, n_features*) sebagai X dan jumlah komponen utama yang ingin diekstraksi sebagai *n_components*. Fungsi ini merupakan implementasi dari PCA (*Principal Component Analysis*) yang akan memberikan hasil berupa matriks data yang sudah tertransformasi.

3. Pembahasan

Pada pengaplikasian metode analisis komponen utama (PCA) kode-kode berikut akan mendemonstrasikan implementasi PCA pada *dataset* iris, yang terdiri dari 150 sampel dari 3 spesies bunga iris dengan 4 fitur.

```
[1] import numpy as np
import pandas as pd
```

Gambar 1. *Import library*

Pada kode “import numpy as np” berfungsi untuk mengimpor modul atau pustaka numpy, biasanya numpy digunakan untuk komputasi statistik. Lalu “import pandas as pd” digunakan untuk mengimpor modul atau pustaka pandas, biasanya pandas digunakan untuk analisis data.

```
[2] def pca(X, n_components):
    X_centered = X - np.mean(X, axis=0)
    U, S, Vt = np.linalg.svd(X_centered)
    return U[:, :n_components] @ np.diag(S[:n_components])
```

Gambar 2. Inisialisasi fungsi pca

Pada kode diatas untuk mengimplementasikan analisis komponen utama dengan numpy. Pada baris pertama mendeklarasikan fungsi *pca()* dengan dua parameter di dalamnya. Pada baris kedua matriks X dikurangi dengan mengurangi rata-rata setiap kolomnya, lalu *np.mean(X, axis=0)* yang berarti menghitung rata-rata setiap kolom, lalu menjadikan matriks setiap kolomnya mempunyai rata-rata 0. Pada baris ketiga fungsi *np.linalg.svd* digunakan untuk mendekomposisi matriks menjadi tiga matriks yang merepresentasikan *Singular Value Decomposition* (SVD) dari matriks tersebut. Lalu pada baris ke empat untuk mengembalikan *dot product*. Hasil dari kedua blok dalam [] adalah untuk representasi data dalam ruang fitur yang direduksi dengan jumlah komponen utama.

```
[3] iris_data = pd.read_csv("iris.csv")
iris_data
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows x 5 columns

Gambar 3. Import dataset

Pembahasan untuk gambar diatas adalah kode yang diberikan pada baris pertama berfungsi untuk membaca *dataset* pada *dataset* iris, kemudian pada baris kedua codenya berfungsi untuk menampilkan *dataset*-nya, dan didapatkan *output* berupa *dataset* iris dari 0 sampai 149 atau 150 baris, yang didalamnya terdapat *sepal_length*, *sepal_width*, *petal_length*, *petal_width*, dan *species*.

```
[4] iris_data.describe()
```

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

Gambar 4. Statistika deskriptif

Pada pembahasan untuk gambar diatas kode yang diberikan untuk menampilkan statistika deskriptifnya berupa *count* (banyak data), *mean* (rata-rata), *std* (standar deviasi), *min* (nilai terkecil), *max* (nilai terbesar), dan lainnya.

```
[5] x = iris_data.select_dtypes(include=['int', 'float'])
    n_components = 2
```

Gambar 5. Inisialisasi parameter fungsi pca

Pada gambar diatas kode pada baris pertama berfungsi untuk memilih kolom kolom dengan tipe data tertentu, dengan parameter int, dan float yang hanya akan disimpan. Lalu pada baris kedua kita mengindikasikan menggunakan 2 komponen utama dalam analisis PCA.

```
[6] X_pca = pca(X, n_components)
    print("Komponen Utama:\n", X_pca)

Komponen Utama:
[[-2.68412563 -0.31939725]
 [-2.71414169  0.17700123]
 [-2.88899057  0.14494943]
 [-2.74534286  0.31829898]
 [-2.72871654 -0.32675451]
 [-2.28085963 -0.74133045]
 [-2.82053775  0.08946138]
 [-2.62614497 -0.16338496]
 [-2.88638273  0.57831175]
 [-2.6727558  0.11377425]
 [-2.50694709 -0.6450689 ]
 [-2.61275523 -0.01472994]
 [-2.78610927  0.235112 ]
 [-3.22380374  0.51139459]
 [-2.64475039 -1.17876464]
 [-2.38603903 -1.33806233]
 [-2.62352788 -0.81067951]
 [-2.64829671 -0.31184914]
 [-2.19982032 -0.87283904]
 [-2.5879864  -0.51356031]
 [-2.31025622 -0.39134594]
 [-2.54370523 -0.43299606]
 [-3.21593942 -0.13346807]
 [-2.30273318 -0.09870885]
 [-2.35575405  0.03728186]
 [-2.50666891  0.14601688]
 [-2.46882007 -0.13095149]
 [-2.56231991 -0.36771886]
 [-2.63953472 -0.31203998]
 [-2.63198939  0.19696122]
```

Gambar 6. Penggunaan fungsi pca

Pada kode ini kita gunakan fungsi *pca()* yang telah kita definisikan sebelumnya untuk melakukan transformasi pada *dataset*, *Output*-nya akan membuat *dataset* dari hasil yang telah direduksi dimensinya menjadi dua komponen utama sesuai dengan hasil transformasi PCA yang telah dilakukan.

4. Kesimpulan

PCA (*Principal Component Analysis*) adalah metode yang berguna untuk melakukan transformasi *dataset* ke ruang dimensi yang lebih rendah sambil mempertahankan sebagian besar informasi yang relevan. Dalam proses PCA, langkah-langkah yang umum dilakukan meliputi normalisasi data, dekomposisi matriks data menggunakan *Singular Value Decomposition* (SVD), dan pemilihan komponen utama yang paling signifikan. Hasil dari PCA adalah *dataset* baru dari hasil reduksi yang memiliki dimensi lebih rendah, dimana

dimensi-dimensi tersebut dipilih berdasarkan kemampuan mereka untuk menjelaskan variasi dalam *dataset* asli. Transformasi ini memungkinkan kita untuk memahami struktur dan pola dalam *dataset* dengan representasi yang lebih sederhana dan informatif, yang dapat digunakan untuk analisis data lebih lanjut, visualisasi, atau aplikasi lainnya. Dengan menggunakan PCA, kita dapat mengurangi kompleksitas data, meningkatkan efisiensi komputasi, dan mempertahankan inti dari informasi yang terkandung dalam *dataset* asli.

5. Daftar Pustaka

- Lovtyaji, H. A. (2022). *IMPLEMENTASI METODE PRINCIPAL COMPONENT ANALYSIS DALAM MENGANALISIS KERAWANAN SAMBARAN PETIR*. PROGRAM STUDI MATEMATIKA FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM. -
- Nikasari, N. L., Dharmawan, K., & Ayu Made Srinadi, I. G. (2017, Januari). ESTIMASI NILAI AVERAGE VALUE AT RISK PADA SAHAM PORTOFOLIO DENGAN MENGGUNAKAN METODE ANALISIS KOMPONEN UTAMA. *E-Jurnal Matematika*, 6(1), 55-64. unud.ac.id. -
- Santosa, Y. P. (2023). KOMBINASI LINIER TARGET DATA UNTUK REGRESI MULTI TARGET MENGGUNAKAN PRINCIPAL COMPONENT ANALYSIS. *Jurnal Teknologi Terpadu*, 9(1), 1-9. journal.nurulfikri.ac.id. -
- Van Delsen, M.S. N., Wattimena, A.Z., & Saputri, S.D. (2017, Desember). PENGGUNAAN METODE ANALISIS KOMPONEN UTAMA UNTUK MEREDUKSI FAKTOR-FAKTOR INFLASI DI KOTA AMBON. *Jurnal Ilmu Matematika dan Terapan*, 11(2), 109-118. unpatti.ac.id. -