

Principal Component Analysis (PCA) dan Factor Analysis (FA) pada Morfologis Mikroorganisme

Widya Louisa¹, Khalila Salma Dzakia Agram², Ulfa Siti Nuraini³

¹Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya, Surabaya, 60231 Indonesia, email: widya.23180@mhs.unesa.ac.id

²Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya, Surabaya, 60231 Indonesia, email: Khalila.23126@mhs.unesa.ac.id

³Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya, Surabaya, 60231 Indonesia, email: ulfanuraini@unesa.ac.id

BAB 1 Pendahuluan

1.1 Latar Belakang

Perkembangan teknologi sequencing DNA, khususnya NGS dan TGS, telah mempercepat identifikasi mikroorganisme. Namun, dalam kasus seperti syok septik, keputusan klinis masih terhambat oleh lamanya proses identifikasi bakteri dan resistensinya. Di sisi lain, resistensi antibiotik kini menjadi krisis global yang menuntut pendekatan diagnostik yang cepat dan efisien. Salah satu pendekatan alternatif adalah analisis morfometrik mikroorganisme berbasis citra, yang memungkinkan estimasi karakteristik resistensi secara tidak langsung.

Analisis multivariat seperti Principal Component Analysis (PCA) dan Factor Analysis (FA) memiliki peran penting dalam mereduksi kompleksitas data, mengekstraksi pola laten, serta memetakan karakteristik bentuk dan ukuran mikroorganisme yang potensial terkait resistensi antibiotik.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, permasalahan yang menjadi fokus dalam penelitian ini adalah:

1. Bagaimana karakteristik morfometrik mikroorganisme dapat digunakan untuk memprediksi resistensi antibiotik dengan cepat?
2. Bagaimana penerapan metode Principal Component Analysis (PCA) dan Factor Analysis (FA) dalam mengekstrak fitur utama dari data morfometrik mikroorganisme?
3. Sejauh mana keakuratan PCA dan FA efektif dalam mengelompokkan

mikroorganisme berdasarkan fitur morfometrik?

1.3 Tujuan Penelitian

Tujuan utama dari penelitian ini adalah untuk:

1. Mengidentifikasi fitur-fitur morfometrik mikroorganisme yang relevan dalam memprediksi resistensi antibiotik secara cepat.
2. Mengevaluasi efektivitas Principal Component Analysis (PCA) dalam mereduksi dimensi data.
3. Menentukan faktor-faktor utama melalui Factor Analysis (FA) yang mampu menjelaskan variasi morfologis mikroorganisme.

BAB 2 Metodologi Penelitian

2.1 Data

Penelitian ini menggunakan dataset

microbes.csv(<https://www.kaggle.com/datasets/sayansh001/microbes-dataset>), yang berisi data morfometrik dari sepuluh jenis mikroorganisme. Data awal terdiri dari 26 kolom, namun dalam penelitian ini, 10 fitur numerik dipilih untuk dianalisis, karena relevan dengan karakterisasi ukuran dan bentuk mikroba.

2.2 Variabel Penelitian

Variabel-variabel yang digunakan dalam penelitian ini mencerminkan sepuluh karakteristik morfometrik utama dari mikroorganisme yang diamati. Solidity mengukur sejauh mana bentuk mikroorganisme memenuhi area konveksnya, memberikan gambaran mengenai kekompakan struktur bentuknya. Sementara itu, eccentricity menggambarkan tingkat keelongasian mikroorganisme, atau seberapa jauh bentuknya menyimpang dari bentuk lingkaran sempurna. Extent menunjukkan rasio antara luas area mikroorganisme dengan luas kotak pembatas (bounding box) yang mengelilinginya, membantu menilai efisiensi ruang yang ditempati mikroorganisme. EulerNumber mencerminkan kompleksitas bentuk dengan menghitung jumlah objek atau lubang dalam struktur mikroorganisme.

Ukuran dari kotak pembatas tersebut juga diamati melalui BoundingBox2, yang menunjukkan lebar horizontal, dan BoundingBox4, yang menunjukkan tinggi vertikal dari kotak yang mengelilingi mikroorganisme. MinorAxisLength mengacu pada panjang sumbu pendek dari bentuk elips yang diasumsikan sesuai dengan bentuk mikroorganisme, berguna untuk memahami dimensi bentuk secara keseluruhan. Perimeter digunakan untuk mengukur keliling atau panjang dari batas luar mikroorganisme, sedangkan ConvexArea memperlihatkan luas area konveks yang melingkupi mikroorganisme secara

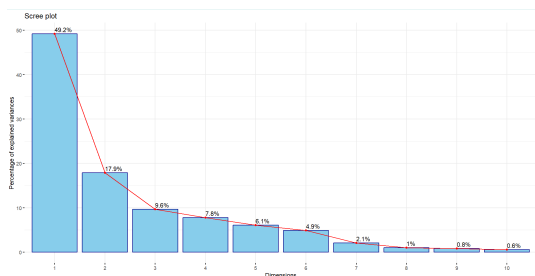
penuhi. Terakhir, Raddi mencerminkan ukuran radius rata-rata yang menggambarkan dimensi menyeluruh dari mikroorganisme secara umum. Keseluruhan variabel ini memberikan fondasi penting dalam menganalisis dan mengkarakterisasi bentuk serta ukuran mikroorganisme secara kuantitatif.

Seluruh data distandardisasi untuk memastikan setiap variabel berkontribusi setara dalam analisis PCA dan FA. Uji KMO dan Bartlett dilakukan untuk menilai kelayakan analisis, dengan hasil KMO sebesar 0,76 dan p-value < 0,05, menandakan data layak untuk PCA-FA.

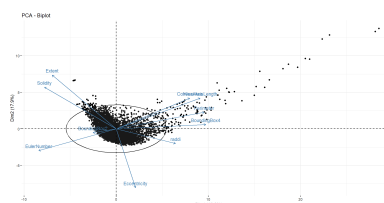
BAB 3 Pembahasan

3.1 Statistik Deskriptif dan Visualisasi

Statistika deskriptif menunjukkan bahwa setiap variabel memiliki rentang nilai yang sangat jauh atau tinggi, seperti *Solidity* dan *Eccentricity* yang bernilai rata-rata tinggi, serta *Perimeter* dan *ConvexArea* yang memiliki nilai maksimum besar meskipun rata-ratanya rendah. Sementara itu, *raddi* menunjukkan nilai yang lebih stabil. Perbedaan skala ini mendorong dilakukannya standarisasi untuk memastikan kontribusi yang seimbang dalam analisis multivariat.



Gambar 3.1



Gambar 3.2

Gambar 3.1, Scree plot hasil analisis Principal Component Analysis (PCA), diketahui bahwa PC1 (49,2%), PC2 (17,9), dan PC3 (9,6%). Jika digabungkan, tiga komponen pertama ini secara total sudah mampu menjelaskan sekitar 76,7% dari keseluruhan variasi yang ada. Dengan mengacu pada scree plot yang dihasilkan, jelas terlihat bahwa grafik menunjukkan penurunan tajam hingga komponen ketiga. Setelah komponen ketiga, grafik mulai melandai, menandakan bahwa tambahan informasi yang diberikan relatif kecil atau tidak signifikan. Berdasarkan pertimbangan ini, dapat disimpulkan bahwa jumlah komponen utama yang optimal untuk menggambarkan data ini adalah sebanyak 3 komponen. Penentuan ini juga diperkuat oleh kriteria nilai eigenvalue yang lebih besar dari 1, di mana hanya tiga komponen pertama yang memenuhi kriteria tersebut, sehingga dianggap cukup untuk dilakukan analisis tanpa kehilangan informasi penting.

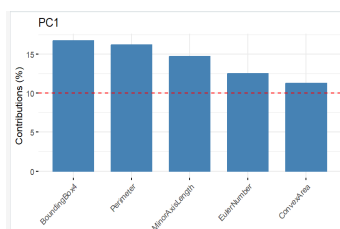
Gambar 3.2, PCA-Biplot memvisualisasikan sebagai vektor berupa panah biru pada biplot PCA, dimana sudut kecil antar vektor menunjukkan korelasi positif kuat, seperti antara variabel "*ConvexArea*" dan "*Perimeter*". Sementara itu, distribusi titik-titik hitam pada biplot merepresentasikan mikroorganisme. Sebagai ilustrasi, titik yang berada jauh dari pusat (kanan biplot) menunjukkan mikroorganisme dengan ukuran besar, sedangkan titik-titik yang berada lebih dekat ke arah vektor "*Extent*" dan "*Solidity*" menggambarkan bentuk yang lebih kompak atau padat. Dengan demikian, biplot ini memperlihatkan bahwa variasi data yang mendominasi adalah ukuran fisik (misalnya, luas area dan panjang perimeter), serta bentuk dan kekompakan mikroorganisme. Hal ini memudahkan dalam mengidentifikasi kelompok berdasarkan karakteristik morfometriknya.

3.2 Asumsi Kelayakan PCA

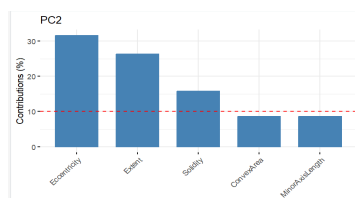
Sebelum menerapkan *Principal Component Analysis* (PCA), variabel perlu distandarisasi menggunakan *z-score* untuk memastikan kontribusi seimbang. Kelayakan PCA diuji melalui dua pendekatan: *Kaiser-Meyer-Olkin* (KMO) sebesar 0,76 menunjukkan bahwa hubungan antar variabel cukup kuat dan *Bartlett's Test of Sphericity* menghasilkan chi-square sebesar 261.212,4 dengan *p-value* < 0,05, mengindikasikan bahwa matriks korelasi signifikan. PCA dengan matriks kovarians menunjukkan tiga komponen utama pertama menjelaskan 76,7% variasi data, dengan kontribusi masing-masing sebesar 49,2% (PC1), 17,9% (PC2), dan 9,6% (PC3). Setelah komponen ketiga, *scree plot* mulai melandai, menunjukkan bahwa informasi penting sebagian besar telah terwakili oleh tiga komponen ini.

3.3 Analisis PCA

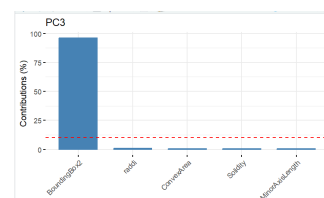
PC1 (49,2%), PC2 (17,9%), dan PC3 (9,65%). Ketiganya mencakup 76,75% informasi total, cukup kuat untuk digunakan sebagai representasi utama data.



Gambar 3.3.1



Gambar 3.3.2

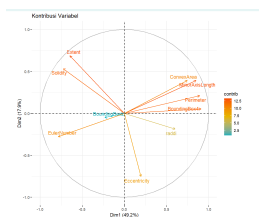


Gambar 3.3.3

Gambar 3.3.1, PC1 (49,22%) total variasi data, dengan variabel BoundingBox4 dan Perimeter menjadi penyumbang tertinggi (>17%), diikuti oleh MinorAxisLength (~15%), serta EulerNumber dan ConvexArea (>10%). Variabel ini mencerminkan dimensi ukuran, bentuk luar, dan kompleksitas objek mikroorganisme. Dengan demikian, PC1 dapat digunakan sebagai indikator umum karakteristik morfometrik mikroba dalam ruang berdimensi rendah.

Gambar 3.3.2, PC2 (17,89%) variasi data dan berfokus pada bentuk mikroorganisme. Variabel Eccentricity menjadi kontributor utama (>30%), Extent (27%) dan Solidity (15%), yang mencerminkan kelonjongan, kepadatan, dan proporsi spasial. Sebaliknya, variabel seperti ConvexArea dan MinorAxisLength memiliki kontribusi rendah (<10%). PC2 melengkapi PC1 dengan merepresentasikan variasi bentuk dan struktur morfologi.

Gambar 3.3.3, PC3 (9,65%) variasi data, didominasi BoundingBox2 (~100%) merepresentasikan panjang sisi horizontal objek mikroorganisme. Variabel lain seperti raddi, ConvexArea, Solidity, dan MinorAxisLength memberikan kontribusi sangat kecil dan dapat diabaikan. Dengan demikian, PC3 mencerminkan fitur spasial unik yang dapat membedakan karakteristik morfometrik mikroba tertentu.

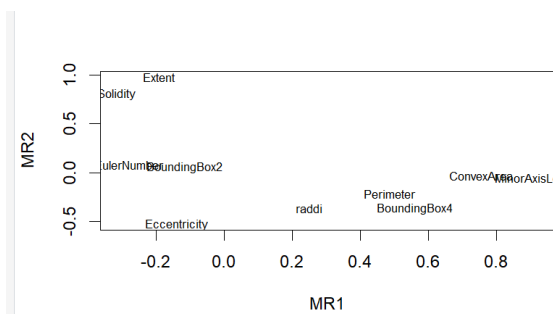


Gambar 3.3.4

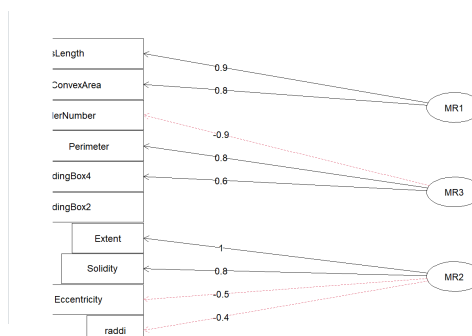
Gambar 3.4.3, menunjukkan kontribusi variabel terhadap dua dimensi utama PCA. Variabel *ConvexArea*, *MinorAxisLength*, *Perimeter*, *BoundingBox4*, *Solidity*, dan *Extent* memiliki kontribusi tinggi (>10), ditunjukkan dengan panah merah-oranye. Sementara *raddi*, *Eccentricity*, *EulerNumber*, dan *BoundingBox2* memiliki kontribusi lebih rendah (<10), ditunjukkan dengan panah kuning hingga biru. Dimensi 1 mendominasi variasi data sebesar 49,2%, diikuti Dimensi 2 sebesar 17,9%.

3.4 Analisis FA

Analisis *Factor Analysis* (FA) dilakukan untuk mengidentifikasi struktur laten dari sepuluh variabel morfometrik mikroorganisme, dengan Principal Axis Factoring dan rotasi Varimax, diperoleh tiga faktor utama (MR1, MR2, dan MR3) yang merangkum sebagian besar informasi dari data yang telah dinormalisasi. Eigenvalues tertinggi dipilih untuk matriks loading yang diputar untuk meningkatkan interpretasi.



Gambar 3.4.1



Gambar 3.4.2

Gambar 3.4.1, menunjukkan hasil *Factor Analysis* yang mengelompokkan variabel morfometrik ke dalam dua faktor utama. *MinorAxisLength*, *ConvexArea*, dan *Perimeter* mendominasi MR1 sebagai representasi ukuran mikroorganisme, sedangkan *Extent* dan *Solidity* mendominasi MR2 sebagai indikator kerapatan bentuk. Sementara *Eccentricity* berkorelasi negatif terhadap MR2. Hasil ini mendukung peran FA dalam mereduksi dimensi dan mengidentifikasi pola morfometrik yang relevan untuk estimasi resistensi antibiotik.

Gambar 3.4.2, menunjukkan struktur tiga faktor utama hasil analisis faktor terhadap sepuluh variabel morfometrik mikroorganisme:

- MR1 berhubungan kuat dengan *MinorAxisLength*, *ConvexArea*, dan *Perimeter*, sehingga diinterpretasikan sebagai Faktor Ukuran.
- MR2 berkorelasi positif dengan *Extent* dan *Solidity*, serta negatif dengan *Eccentricity* dan *raddi*, mencerminkan Faktor Kerapatan dan Simetri.
- MR3 menunjukkan korelasi negatif kuat dengan *EulerNumber* dan positif dengan *BoundingBox4*, merepresentasikan Faktor Kompleksitas Bentuk.

Ketiga faktor ini membantu mereduksi dimensi dan menyederhanakan interpretasi karakteristik morfologis mikroorganisme.

BAB 4 Kesimpulan

Penelitian ini menunjukkan bahwa metode Principal Component Analysis (PCA) dan Factor Analysis (FA) efektif dalam mereduksi kompleksitas data morfometrik mikroorganisme menjadi beberapa dimensi utama yang informatif. PCA berhasil merangkum 76,75% variasi data ke dalam tiga komponen utama, dengan PC1 mewakili ukuran, PC2 mencerminkan bentuk dan kepadatan, serta PC3 merepresentasikan aspek spasial. FA memperkuat temuan ini dengan tiga faktor laten: ukuran (MR1), kerapatan dan simetri (MR2), serta kompleksitas bentuk (MR3).

Hasil ini selaras dengan tujuan penelitian, yaitu mengekstraksi fitur penting dari bentuk mikroorganisme untuk mendukung estimasi resistensi antibiotik secara cepat melalui citra. Pendekatan ini berpotensi digunakan sebagai metode awal dalam sistem diagnostik cepat, khususnya pada kondisi kritis seperti syok septik, sebelum hasil uji resistensi molekuler tersedia.

Daftar Pustaka

Afzal, A., Khan, M. A., Sharif, M., Shad, S. A., & Kim, K. W. (2021). Image-based identification of antimicrobial resistant bacteria using deep learning and feature reduction techniques. *Computers in Biology and Medicine*, 139, 104931.

<https://doi.org/10.1016/j.compbimed.2021.104931>

- Ba, I., Turgeon, M., Veniamin, S., Joel, J., Miller, R., Graham, M., & McGregor, K. (2024). A Bayesian factor analysis model for high-dimensional microbiome count data. *arXiv*.
<https://arxiv.org/abs/2404.03127>
- Carabidus. (2019). *Principal Component Analysis (PCA) in R*. RPubS.
<https://rpubs.com/carabidus/465971>
- CD Genomics. (n.d.). *PCA and Principal Coordinate Analysis (PCoA) for microbial sequencing: Introduction and procedures*.
<https://www.cd-genomics.com/microbioseq/principal-component-analysis-pca-and-principal-coordinate-analysis-pcoa-for-microbial-sequencing-introduction-and-procedures.html>
- Everitt, B. S., & Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. Springer.
- Gao, C., & Wang, Y. (2022). Microbial community assembly and interactions in extreme environments. *Environmental Microbiology*, 24(5), 2003–2015.
<https://doi.org/10.1111/1462-2920.16218>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson Education.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Namira_Romadhona. (2023). *PCA kualitas air*. RPubS.
https://rpubs.com/namira_romadhona/pcakualitasair
- Ramos, M., Coelho, F., Oliveira, A., & Silva, L. (2020). Application of multivariate statistical methods for bacterial classification based on morphological features. *Journal of Microbiological Methods*, 170, 105850. <https://doi.org/10.1016/j.mimet.2020.105850>
- RPubs. (n.d.). *PCA & FA untuk data biomedis*.
<https://rpubs.com/sinhrks/pca>
https://rpubs.com/ajabri/factor_analysis
- Satrio0126. (2023). *PCA analysis on dataset 1111592*. RPubS.
<https://rpubs.com/satrio0126/1111592>
- Shifaaainun. (2021). *PCA on healthy lifestyle data*. RPubS.
<https://rpubs.com/Shifaaainun/PCAHealthyLifestyle>
- USDA Forest Service. (n.d.). *Principal component analysis of microbial community data from*

an experimental decay study. <https://research.fs.usda.gov/treesearch/46268>

World Health Organization (WHO). (2020). *Antimicrobial resistance*.

<https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>