



Tinjauan Komprehensif tentang Aplikasi dan Perkembangan *Principal Component Analysis* (PCA)

Faris Alaudin Shalih^{1*}, Rafli Akbar Ramadhan¹, Najlaa Syalaisa¹

¹Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Militer,
Universitas Pertahanan Republik Indonesia, Bogor, Indonesia

*Correspondence author: falaudin18@gmail.com

ABSTRAK

Principal Component Analysis (PCA) adalah metode statistik yang populer untuk mereduksi dimensi data dalam analisis data kompleks. Dikembangkan sejak awal abad ke-20 oleh Pearson dan Hotelling, PCA telah menjadi alat penting di berbagai disiplin ilmu. Artikel ini mengulas penerapan PCA di berbagai industri untuk menjelaskan prinsip kerja dan langkah-langkah utama dalam metode ini, seperti normalisasi data, perhitungan matriks kovarians, dekomposisi nilai eigen, dan proyeksi data. Ulasan juga mencakup studi-studi terdahulu yang menerapkan PCA di bidang kesehatan, penginderaan jauh, finansial, pertanian, pendidikan, dan geologi. Hasilnya menunjukkan bahwa PCA efektif dalam mengurangi dimensi data, meningkatkan akurasi, dan mengidentifikasi fitur signifikan pada setiap aplikasi. PCA juga berkontribusi pada pengambilan keputusan berbasis data yang akurat. Hingga saat ini, PCA tetap aplikatif dan memiliki penerapan yang luas di berbagai bidang, membuktikan efektivitasnya pada data modern yang terus berkembang.

© 2025 Kantor Jurnal dan Publikasi UPI

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima 9 November 2024
Direvisi 1 Desember 2024
Disetujui 1 April 2025
Tersedia online 2 Mei 2025
Dipublikasikan 2 Mei 2025

Kata Kunci:

Aplikasi PCA,
Linimasa PCA,
Metodologi PCA,
Principal Component Analysis (PCA),
Reduksi Dimensi.

ABSTRACT

Principal Component Analysis (PCA) is a popular statistical method for reducing data dimensionality in complex data analysis. Developed since the early 20th century by Pearson and Hotelling, PCA has become an important tool in various disciplines. This article reviews the application of PCA in various industries to explain the working principles and key steps in this method, such as data normalization, covariance matrix calculation, eigenvalue decomposition, and data projection. The review also includes previous studies applying PCA in healthcare, remote sensing, finance, agriculture, education, and geology. The results show that PCA effectively reduces data dimensionality, improves accuracy, and identifies significant features in each application. PCA also contributes to accurate data-driven decision-making. To date, PCA remains applicable and has wide applicability in various fields, proving its effectiveness in facing the growing challenges of modern data.

Keywords:

Analysis (PCA),
Dimensionality
Reduction,
PCA Applications,
PCA Methodology,
PCA Timeline,
Principal Component.

1. PENDAHULUAN

Penerapan analisis data dalam berbagai ilmu telah berkembang pesat seiring dengan kompleksitas data yang ada (Gewers et al., 2021). Salah satu metode untuk mengatasinya yaitu Principal Component Analysis (PCA). PCA merupakan salah satu teknik analisis statistik yang paling umum digunakan untuk mereduksi dimensi data (Reddy et al., 2020; Spiegelberg & Rusz, 2017; Weylandt & Swiler, 2024; Zou & Xue, 2018). Metode ini pertama kali diperkenalkan pada awal abad ke-20 dan terus berkembang seiring dengan meningkatnya kompleksitas data yang dihadapi oleh para peneliti dan praktisi di berbagai bidang. Berbagai penelitian menunjukkan bahwa (AbdelFattah et al., 2021; Dharmawan et al., 2023; Ding & Miao, 2020; Gewers et al., 2021; Jakaitiene et al., 2018; Kartika et al., 2023; Olle et al., 2024). Pada tahun 1901, Karl Pearson pertama kali memperkenalkan konsep PCA sebagai metode untuk menemukan "garis terbaik" yang meminimalkan kesalahan kuadrat dalam data multidimensi (Pearson, 1901). Pada tahun 1933, Harold Hotelling mengembangkan ide Pearson dengan mengusulkan PCA sebagai metode transformasi variabel berkorelasi menjadi sekumpulan variabel yang tidak berkorelasi (komponen utama), yang dikenal sebagai transformasi ortogonal. Kontribusi Hotelling adalah memformalkan dasar matematika PCA dan memperkenalkan konsep transformasi ortogonal.

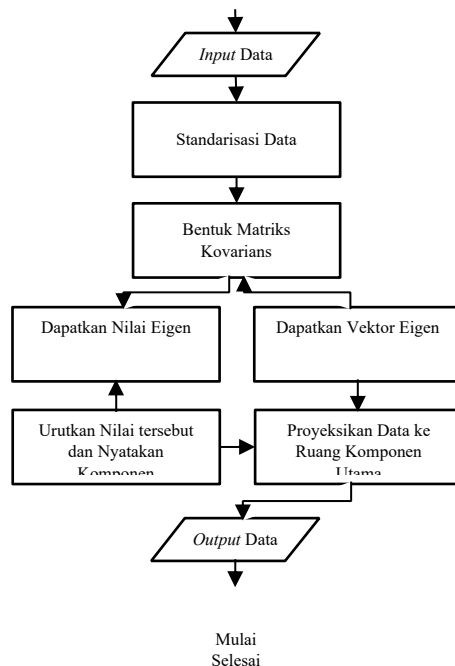
Tahun 1967, Gene H. Golub dan William Kahan memperkenalkan metode komputasi efisien untuk PCA menggunakan Singular Value Decomposition (SVD). SVD memungkinkan perhitungan komponen utama yang lebih efisien, sehingga PCA menjadi lebih praktis untuk analisis data besar (Zheng et al., 2023). Pada tahun 1986, Ian Jolliffe mempopulerkan penggunaan PCA dalam berbagai disiplin ilmu melalui bukunya "*Principal Component Analysis*," yang menjadi referensi penting bagi para praktisi. Principal Component Analysis (PCA) adalah metode statistik yang umum digunakan untuk mereduksi dimensi data dan menangani kompleksitas data (Cardot & Degras, 2018; Gewers et al., 2022; Lever et al., 2017; Maadooliat et al., 2015). Sejak diperkenalkan oleh Karl Pearson pada tahun 1901, PCA telah menjadi alat esensial dalam berbagai bidang, mulai dari pengolahan citra hingga analisis genetik (Bharadiya, 2023; Iqbal & Kumar, 2022). Saat ini, PCA banyak diterapkan dalam machine learning dan big data karena kemampuannya mereduksi kompleksitas data tanpa menghilangkan informasi penting (Bharadiya, 2023; Ray et al., 2020). Dalam analisis data multidimensi, PCA berperan penting dalam mengidentifikasi pola tersembunyi dan mengurangi noise (Alessio & Cannistraci, 2016; Jackle et al., 2016). Metode ini juga membantu mencegah overfitting dalam model pembelajaran mesin yang beroperasi pada data berdimensi tinggi (Dong et al., 2023; Raiko et al., 2007; Rubaie et al., 2017). Penerapannya yang luas dan fleksibel membuat PCA menjadi salah satu metode pilihan utama dalam banyak aplikasi analisis data, mulai dari pengolahan citra dan sinyal, hingga analisis genetik dan keuangan (AbdelFattah et al., 2021; Dharmawan et al., 2023; Ding & Miao, 2020; Jakaitiene et al., 2018; Kartika et al., 2023; Olle et al., 2024).

Seiring dengan perkembangan teknologi dan kompleksitas data, penerapan PCA terus berkembang. Penelitian terbaru menunjukkan variasi dan modifikasi metode PCA yang lebih adaptif terhadap data modern, terutama di bidang kecerdasan buatan dan big data (Halko et al., 2011; Jianqing et al., 2019). Meskipun banyak studi tentang PCA, masih ada kesenjangan

pengetahuan terkait aplikasi spesifik dan tantangan teknis dalam penerapannya di berbagai industri. Artikel review ini disusun untuk menjawab permasalahan tersebut dengan memberikan pandangan menyeluruh tentang bagaimana PCA saat ini digunakan di berbagai industri dan ilmu pengetahuan. Kebaruan dari artikel ini adalah pengulasan dari penerapan PCA pada berbagai industri dan ilmu pengetahuan. Artikel ini bertujuan memberikan tinjauan menyeluruh tentang perkembangan PCA serta aplikasinya yang terkini. Teknis dari review ini adalah menggabungkan beberapa sumber literatur yang relevan, dimana harapannya artikel ini menjadi referensi berguna bagi akademisi dan praktisi dalam memahami penggunaan PCA dalam konteks modern serta potensi pengembangannya di masa depan.

2. METODE

Artikel ini akan memberikan gambaran menyeluruh mengenai penerapan metode PCA di berbagai industri guna memahami secara umum prinsip kerja dari metode PCA. Dasar penggunaan metode ini menurut (Bloemendal et al., 2016; Erdogmus et al., 2004; Fan et al., 2021; Wilks, 2011) akan dijelaskan melalui Gambar 1.



Gambar 1. Flowchart metode PCA secara umum

Berdasarkan Gambar 1, dapat dijabarkan sebagai berikut:

Tahapan dimulai dengan normalisasi data untuk memastikan setiap fitur memiliki rata-rata 0 dan standar deviasi 1, sehingga fitur yang memiliki skala berbeda tidak mendominasi komponen utama yang akan dihasilkan. Normalisasi data dirumuskan sebagai Persamaan 1 berikut .

$$x_{norm} = \frac{x - \mu}{\sigma} \quad (1)$$

Di mana x adalah nilai dari suatu fitur, μ adalah rata-rata dari fitur tersebut, dan σ adalah standar deviasi dari fitur tersebut. Setelah itu, matriks kovarians dihitung untuk menilai hubungan linier antara fitur-fitur dalam dataset dengan Persamaan 2 berikut.

$$Cov(X) = \frac{1}{n-1}(X^T X) \quad (2)$$

Di mana X adalah matriks data yang sudah dinormalisasi dan n adalah jumlah observasi. Berdasarkan matriks kovarians ini, nilai eigen (λ) dan vektor eigen (μ) dihitung melalui Persamaan 3 berikut.

$$Cov(X)\mu = \lambda\mu \quad (3)$$

Nilai eigen mengindikasikan besarnya variansi data sepanjang arah vektor eigen yang bersangkutan.

Komponen utama kemudian dipilih berdasarkan nilai eigen terbesar, yang menunjukkan arah variansi maksimum dalam data. Pada beberapa kasus, hanya terdapat beberapa komponen utama dengan nilai eigen terbesar yang dipertahankan untuk mengurangi dimensi data tanpa kehilangan banyak informasi. Selanjutnya data asli diproyeksikan ke ruang komponen utama menggunakan vektor eigen yang telah dipilih, hal ini akan menghasilkan data berdimensi lebih rendah namun tetap mengandung informasi yang signifikan, dengan persamaan transformasi pada Persamaan 4 berikut.

$$Z = X \cdot V \quad (4)$$

Di mana Z adalah data yang diproyeksikan, X adalah data asli, dan V adalah matriks vektor eigen yang dipilih. Selanjutnya sebagai upaya melihat gambaran menyeluruh mengenai penerapan metode PCA di berbagai industri, kami telah mengumpulkan beberapa literatur terkait penggunaan PCA yang telah dikembangkan sedemikian rupa untuk penerapan beberapa bidang di industri. Penerapan PCA dapat telah dihimpun dalam Tabel 1 berikut.

Tabel 1. Pekerjaan Sebelumnya Tentang Penerapan PCA di Beberapa Bidang

No.	Judul Studi	Bidang	Ref.
1.	<i>Analysis of Education Systems Performance in European Countries by Means of PCA-DEA</i>	Pendidikan	(Jakaitiene et al., 2018)
2.	<i>Collapsed Building Classification with Optical and SAR Data Based on Manifold Learning</i>	Penginderaan Jauh	(Ding & Miao, 2020)
3.	<i>Quantitative Evaluation of Soil Quality Using Principal Component Analysis: The Case Study of El-Fayoum Depression Egypt</i>	Geologi	(AbdelFattah et al., 2021)
4.	Implementasi Algoritma <i>Principal Component Analysis</i> (PCA) dan <i>KNearest Neighbor</i> (KNN) untuk Memprediksi Kelayakan Kredit Pengguna Smartphone di Indonesia pada Masa Pandemi Covid-19	Finansial	(Kartika et al., 2023)
5.	<i>Development of PCA-MLP Model Based on Visible and Shortwave Near Infrared Spectroscopy for Authenticating Arabica Coffee Origins</i>	Agrikultur	(Dharmawan et al., 2023)

- | | | | |
|----|--|-------|---------------------|
| 6. | <i>Application and comparison of K-means and PCA-based segmentation models for Alzheimer's disease detection using MRI</i> | Medis | (Olle et al., 2024) |
|----|--|-------|---------------------|
-

3. HASIL DAN PEMBAHASAN

Berdasarkan artikel pertama yang telah direview sebelumnya, artikel ini merupakan salah satu contoh penerapan PCA pada bidang pendidikan (Jakaitiene et al., 2018). Khususnya analisis sistem pendidikan di negara-negara Eropa. Metode dalam artikel ini menerapkan PCA untuk mereduksi kompleksitas data dari delapan indikator pendidikan menjadi empat komponen utama yang menangkap 90% variabilitas data. Selanjutnya dikombinasikan dengan Data Envelopment Analysis (DEA) untuk mengukur efisiensi relatif setiap negara. Hasil yang diperoleh menunjukkan bahwa beberapa negara seperti Denmark, Irlandia, dan Swiss mencapai frontier efisiensi, sedangkan negara-negara seperti Bulgaria dan Latvia memiliki jarak signifikan dari frontier efisiensi, sehingga membutuhkan perbaikan dalam indikator kinerja pendidikan mereka.

Artikel kedua yang direview merupakan contoh penerapan PCA dalam bidang penginderaan jauh (Ding & Miao, 2020). Khususnya untuk klasifikasi bangunan runtuh pasca bencana menggunakan data optik dan SAR. Penelitian ini menerapkan PCA sebagai metode reduksi dimensi untuk memproyeksikan data dari citra polarimetrik RADARSAT-2 dan citra resolusi tinggi Google Earth ke ruang fitur berdimensi lebih rendah. Pola distribusi dari enam kelas CBC (tanah, vegetasi jarang, air, batu, bangunan runtuh, dan bangunan utuh) dapat diidentifikasi dan divisualisasikan dengan lebih baik. Hasilnya menunjukkan bahwa PCA memberikan akurasi klasifikasi sebesar 97.13% dengan nilai Kappa 0.9639, lebih tinggi dibandingkan metode lain seperti ISOMAP dan LLE. PCA terbukti efektif dalam memisahkan kelas-kelas CBC, terutama untuk kelas-kelas yang sulit dibedakan seperti tanah dan bangunan runtuh, menjadikannya metode yang unggul dalam analisis citra satelit untuk aplikasi penginderaan jauh pasca bencana.

Artikel ketiga memberikan gambaran penerapan PCA pada bidang geologi (AbdelFattah et al., 2021). Khususnya evaluasi kualitas tanah. Metode dalam artikel ini menerapkan PCA untuk mereduksi kompleksitas data dari 36 sampel tanah yang dikumpulkan di El-Fayoum, Mesir. Hal ini dilakukan dengan tujuan menghindari multikolinearitas di antara variabel fisik dan kimia tanah seperti pH, salinitas, kapasitas tukar kation, dan kandungan nitrogen. PCA digunakan untuk mengidentifikasi komponen utama yang paling berpengaruh terhadap kualitas tanah dan menghitung bobot relatif dari indikator tersebut guna membentuk Soil Quality Index (SQI). Data tersebut kemudian dimasukkan ke dalam model statistik kriging biasa dalam ArcGIS untuk membuat peta distribusi spasial kualitas tanah. Hasil yang diperoleh menunjukkan bahwa tiga komponen utama (PC1, PC2, dan PC3) menjelaskan 83,63% variansi data, dengan PC1 memberikan pengaruh terbesar terkait kapasitas tukar kation, salinitas, dan bahan organik tanah. Peta distribusi spasial SQI membagi wilayah penelitian menjadi tiga kategori kualitas tanah, yaitu sangat baik, baik, dan cukup baik, dengan 14,48% area berada dalam kategori sangat baik. Penelitian ini berhasil mereduksi data kompleks dan memberikan alat visual penting untuk pengelolaan tanah presisi di El-Fayoum, membantu optimalisasi lahan pertanian di wilayah tersebut.

Artikel keempat yang direview memberikan gambaran penerapan PCA dan KNN dalam bidang finansial (Kartika et al., 2023). Khususnya untuk memprediksi kelayakan kredit pengguna smartphone di Indonesia selama pandemi. Metode yang digunakan melibatkan PCA untuk mereduksi dimensi data yang diperoleh dari kuesioner online yang dikumpulkan dari 1050 responden. Hal ini dilakukan dengan tujuan mendapatkan variabel yang paling berpengaruh dalam prediksi kelayakan kredit. Alhasil dari 12 variabel, 9 yang paling signifikan dipilih, sementara 3 variabel lainnya diabaikan. Dalam teknisnya, setelah pengurangan dimensi oleh PCA, data diolah dengan algoritma K-Means untuk mengidentifikasi kluster-kluster yang relevan sebelum diterapkan K-Nearest Neighbor (KNN) untuk klasifikasi akhir. Hasil penelitian menunjukkan bahwa model KNN memberikan akurasi 85%, dengan presisi 87%, recall 84%, dan F1-score 85%, serta mampu mengidentifikasi tiga kluster pengguna berdasarkan perilaku penggunaan smartphone dan variabel demografis. Studi ini juga menyimpulkan bahwa data penggunaan smartphone dapat menjadi alternatif yang kuat untuk penilaian kredit bagi individu yang tidak memiliki sejarah keuangan formal, yang menunjukkan potensi besar dalam proses penilaian kredit di masa depan.

Artikel kelima yang direview merupakan contoh penerapan Principal Component Analysis (PCA) dalam bidang pertanian (Dharmawan et al., 2023). Khususnya untuk autentikasi asal geografis kopi Arabika. Metode yang digunakan dalam penelitian ini melibatkan pengumpulan data spektroskopi dari biji kopi Arabika yang berasal dari empat daerah di Indonesia, yaitu Temanggung, Toraja, Gayo, dan Kintamani. Pertama dilakukan proses preprocessing data untuk menghilangkan noise dan efek scattering. PCA digunakan untuk mereduksi dimensi data dan mengidentifikasi pola distribusi berdasarkan asal kopi. Dua komponen utama pertama dari PCA berhasil menjelaskan lebih dari 90% varians dalam data. Data yang dihasilkan kemudian digunakan sebagai input untuk model Multilayer Perceptron (MLP), yang dilatih untuk mengklasifikasikan asal geografis kopi dengan arsitektur jaringan yang terdiri dari dua lapisan tersembunyi dan satu lapisan output. Evaluasi model dilakukan dengan metode k-fold cross-validation (k=10). Metode ini berhasil menunjukkan akurasi tinggi antara 90% hingga 100% dalam mengidentifikasi asal geografis kopi, dengan nilai spesifisitas, presisi, recall, dan F-score yang tinggi.

Artikel keenam merupakan salah satu contoh penerapan PCA dalam bidang Kesehatan (Olle et al., 2024). Khususnya untuk deteksi dini penyakit Alzheimer menggunakan citra MRI. Metode yang digunakan dalam penelitian ini melibatkan penerapan PCA untuk melakukan reduksi dimensi citra MRI berukuran 256x256 piksel. Setiap piksel bertindak sebagai variabel yang direduksi menjadi 50 komponen utama, yang tetap mempertahankan sekitar 90% dari variansi data asli. Penulis mengekstraksi fitur dari area penting otak seperti materi abu-abu, materi putih, dan cairan serebrospinal, yang relevan dalam analisis penyakit Alzheimer karena perubahan degeneratif pada materi abu-abu. Setelah proses reduksi dimensi dengan PCA, data ini digunakan untuk melatih model PCA yang digabungkan dengan ANN (Artificial Neural Network). Dalam hasilnya, model PCA+ANN terbukti lebih unggul dibandingkan model k-means clustering yang digabungkan dengan CNN (Convolutional Neural Network). Tingkat akurasi model ini mencapai 91% dan AUC 0.941, lebih tinggi dibandingkan dengan k-means+CNN yang hanya mencapai akurasi 82% dan AUC 0.887. Selain itu, PCA+ANN juga lebih efisien dalam waktu pemrosesan, hanya memerlukan 106.70 detik untuk 500 epoch, sedangkan model k-means+CNN membutuhkan 11.50 jam untuk 60 epoch.

Penerapan *Principal Component Analysis* (PCA) dalam berbagai bidang menunjukkan efektivitasnya sebagai alat analisis yang kuat dan fleksibel (Lever et al., 2017; Zou & Xue, 2018). Hal ini memberikan kontribusi signifikan dalam penelitian yang berfokus pada pengolahan data yang kompleks (Lever et al., 2017; Tsagkarakis et al., 2018; Wu et al., 2018). Dalam bidang kesehatan, PCA berhasil mengurangi dimensi citra MRI untuk deteksi penyakit Alzheimer, yang meningkatkan akurasi model dengan mengidentifikasi fitur penting yang relevan terhadap kondisi degeneratif otak. Dalam bidang penginderaan jauh, PCA memfasilitasi klasifikasi bangunan runtuh pasca bencana, membedakan kelas yang sulit dibedakan dengan akurasi tinggi, sehingga mendukung pemulihan dan manajemen bencana. Sementara itu, dalam bidang finansial, PCA mendukung analisis kelayakan kredit dengan mereduksi variabel dari kuesioner menjadi indikator yang paling berpengaruh. Hal ini memungkinkan identifikasi kluster pengguna yang lebih tepat. Penggunaan bidang pertanian, PCA membantu dalam autentikasi asal kopi Arabika berdasarkan data spektroskopi, menunjukkan kemampuannya dalam mengidentifikasi pola distribusi yang berhubungan dengan kualitas produk. Dalam bidang pendidikan, PCA berperan dalam menyaring informasi penting untuk menilai efisiensi, memberikan rekomendasi untuk perbaikan. Bidang geologi, PCA terbukti efektif dengan kemampuannya untuk mereduksi kompleksitas data dan mengidentifikasi faktor-faktor utama yang berkontribusi terhadap variabilitas tanah. PCA tidak hanya membantu menghindari multikolinearitas antar variabel, tetapi juga memungkinkan pembuatan peta distribusi spasial yang memberikan wawasan visual yang penting bagi pengelolaan tanah presisi, mendukung upaya optimalisasi lahan pertanian. Secara keseluruhan, PCA terbukti efektif dalam mereduksi kompleksitas data, meningkatkan pemahaman, dan menghasilkan kebijakan berbasis data yang lebih akurat di berbagai disiplin ilmu, menunjukkan potensi luasnya untuk aplikasi lebih lanjut dalam penelitian-penelitian masa depan.

4. KESIMPULAN

Metode *Principal Component Analysis* (PCA) telah terbukti efektif dan luas diterapkan dalam berbagai bidang. Penerapan PCA memungkinkan peneliti untuk mereduksi kompleksitas data, mengidentifikasi komponen utama yang signifikan, dan menghasilkan informasi yang lebih terfokus dan dapat diandalkan untuk pengambilan keputusan. Dalam setiap contoh penelitian yang diulas, PCA tidak hanya berfungsi sebagai alat analisis, tetapi juga memberikan wawasan berharga yang dapat diterapkan dalam praktik nyata. Meskipun banyak kemajuan telah dicapai melalui penggunaan PCA, terdapat kebutuhan mendesak untuk mengeksplorasi lebih dalam penerapan metode ini di berbagai bidang yang berkembang pesat. Penyusunan artikel review ini diharapkan dapat memberikan pemahaman yang komprehensif mengenai cara kerja PCA, manfaatnya, serta aplikasinya dalam konteks modern. Dengan pemahaman yang lebih baik mengenai PCA, baik akademisi maupun praktisi dapat memanfaatkan metode ini untuk meningkatkan penelitian dan inovasi di bidang mereka masing-masing, sehingga mendorong pengembangan ilmu pengetahuan dan teknologi yang lebih lanjut.

Saran dari penulis, Perlu dilakukan penelitian lanjutan yang mengeksplorasi potensi dan batasan dari PCA dalam konteks yang lebih luas, termasuk pengembangan metode baru yang dapat mengatasi tantangan yang mungkin muncul saat menerapkan PCA pada dataset yang besar dan kompleks. Selain itu, kolaborasi antara peneliti dari berbagai disiplin ilmu

dapat mendorong inovasi dan penerapan PCA yang lebih efektif, memastikan bahwa metode ini tetap relevan dan bermanfaat di masa depan.

5. REFERENSI

- AbdelFattah, M. K., Mohamed, E. S., Wagdi, E. M., Shahin, S. A., Aldosari, A. A., Lasaponara, R., & Alnaimy, M. A. (2021). Quantitative evaluation of soil quality using principal component analysis: The case study of El-Fayoum depression Egypt. *Sustainability (Switzerland)*, *13*(4), 1–19.
- Alessio, M., & Cannistraci, C. V. (2016). *Nonlinear Dimensionality Reduction by Minimum Curvilinearity for Unsupervised Discovery of Patterns in Multidimensional Proteomic Data* (pp. 289–298).
- Bharadiya, J. P. (2023). A tutorial on principal component analysis for dimensionality reduction in machine learning. *International journal of innovative science and research technology*, *8*(5), 2028-2032.
- Bloemendal, A., Knowles, A., Yau, H.-T., & Yin, J. (2016). On the principal components of sample covariance matrices. *Probability Theory and Related Fields*, *164*(1–2), 459–552.
- Cardot, H., & Degras, D. (2018). Online Principal Component Analysis in High Dimension: Which Algorithm to Choose? *International Statistical Review*, *86*(1), 29–50.
- Dharmawan, A., Masithoh, R. E., & Amanah, H. Z. (2023). Development of PCA-MLP model based on visible and shortwave near infrared spectroscopy for authenticating Arabica coffee origins. *Foods*, *12*(11), 2112.
- Ding, L., & Miao, H. (2020). Collapsed building classification with optical and sar data based on manifold learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *5*(3), 83–88.
- Dong, W., Woźniak, M., Wu, J., Li, W., & Bai, Z. (2023). Denoising Aggregation of Graph Neural Networks by Using Principal Component Analysis. *IEEE Transactions on Industrial Informatics*, *19*(3), 2385–2394.
- Erdogmus, D., Rao, Y. N., Peddaneni, H., Hegde, A., & Principe, J. C. (2004). Recursive Principal Components Analysis Using Eigenvector Matrix Perturbation. *EURASIP Journal on Advances in Signal Processing*, *2004*(13), 263984.
- Fan, Z., Sun, Y., & Wang, Z. (2021). Principal components in linear mixed models with general bulk. *The Annals of Statistics*, *49*(3), 1489-1513.
- Gewers, F. L., Ferreira, G. R., Arruda, H. F. De, Silva, F. N., Comin, C. H., Amancio, D. R., & Costa, L. D. F. (2022). Principal Component Analysis. *ACM Computing Surveys*, *54*(4), 1–34.
- Gewers, F. L., Ferreira, G. R., De Arruda, H. F., Silva, F. N., Comin, C. H., Amancio, D. R., & Costa, L. D. F. (2021). Principal component analysis: A natural approach to data exploration. *ACM Computing Surveys*, *54*(4), 1–33.

- Halko, N., Martinsson, P.-G., Shkolnisky, Y., & Tygert, M. (2011). An Algorithm for the Principal Component Analysis of Large Data Sets. *SIAM Journal on Scientific Computing*, 33(5), 2580–2594.
- Iqbal, N., & Kumar, P. (2022). From Data Science to Bioscience: Emerging era of bioinformatics applications, tools and challenges. *Procedia Computer Science*, 218(2022), 1516–1528.
- Jackle, D., Fischer, F., Schreck, T., & Keim, D. A. (2016). Temporal MDS Plots for Analysis of Multivariate Data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 141–150.
- Jakaitiene, A., Zilinskas, A., & Stumbriene, D. (2018). Analysis of education systems performance in European Countries by Means of PCA-DEA. *Informatics in Education*, 17(2), 245–263.
- Fan, J., Wang, D., Wang, K., & Zhu, Z. (2019). Distributed estimation of principal eigenspaces. *Annals of statistics*, 47(6), 3009.
- Kartika, R., Winahyu, K., & Saputro, V. A. (2023). Implementasi Algoritma Principal Component Analysis (PCA) dan K-Nearest Neighbor (KNN) Untuk Memprediksi Kelayakan Kredit Pengguna Smartphone Di Indonesia Pada Masa Pandemi Covid-19. *Jurnal Darma Agung*, 1–10.
- Lever, J., Krzywinski, M., & Altman, N. (2017). Principal component analysis. *Nature Methods*, 14(7), 641–642.
- Maadooliat, M., Huang, J. Z., & Hu, J. (2015). Integrating Data Transformation in Principal Components Analysis. *Journal of Computational and Graphical Statistics*, 24(1), 84–103.
- Olle Olle, D. G., Zoobo Bisse, J., & Abessolo Alo’o, G. (2024). Application and comparison of K-means and PCA based segmentation models for Alzheimer disease detection using MRI. *Discover Artificial Intelligence*, 4(1), 11.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572.
- Raiko, T., Ilin, A., & Karhunen, J. (2007). Principal Component Analysis for Large Scale Problems with Lots of Missing Values. In *Machine Learning: ECML* (pp. 691–698). Springer Berlin Heidelberg.
- Ray, S., AlGhamdi, A., Alshouiliy, K., & Agrawal, D. P. (2020, June). Selecting features for breast cancer analysis and prediction. In *2020 International Conference on Advances in Computing and Communication Engineering (ICACCE)* (pp. 1-6). IEEE.
- Reddy, G. T., Reddy, M. P. K., Lakshmana, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access*, 8, 54776–54788.

- Rubaie, M., Wu, P., Chang, J. M., & Kung, S.-Y. (2017). Privacy-preserving PCA on horizontally-partitioned data. *2017 IEEE Conference on Dependable and Secure Computing*, 280–287.
- Spiegelberg, J., & Rusz, J. (2017). Can we use PCA to detect small signals in noisy data? *Ultramicroscopy*, 172, 40–46.
- Tsagkarakis, N., Markopoulos, P. P., Sklivanitis, G., & Pados, D. A. (2018). L1-Norm Principal-Component Analysis of Complex Data. *IEEE Transactions on Signal Processing*, 66(12), 3256–3267.
- Weylandt, M., & Swiler, L. P. (2024). Beyond PCA: Additional Dimension Reduction Techniques to Consider in the Development of Climate Fingerprints. *Journal of Climate*, 37(5), 1723–1735.
- Wilks, D. S. (2011). Principal component (EOF) analysis. In *International geophysics* (Vol. 100, pp. 519-562). Academic press.
- Wu, S. X., Wai, H.-T., Li, L., & Scaglione, A. (2018). A Review of Distributed Algorithms for Principal Component Analysis. *Proceedings of the IEEE*, 106(8), 1321–1340.
- Zheng, A. Y., Chen, C., Chen, Y. Q., Zhao, B. B., & Zhao, P. D. (2023). Application of SVD combined with PCA in delineation and evaluation of ore-prospecting targets in the Gejiu tin polymetallic cluster region, SW China. *Ore Geology Reviews*, 160, 105571.
- Zou, H., & Xue, L. (2018). A Selective Overview of Sparse Principal Component Analysis. *Proceedings of the IEEE*, 106(8), 1311–1320.