

**Model Kombinasi Optimalisasi PCA dengan Teknik Adaptive Weighting  
untuk Deteksi Anomali pada Dataset Multivariat**



**Dosen : Dr. Basuki Rahmat, S.Si., MT**

oleh :

Alya' Izzah Zalfa' Rihadah Ramadhani Nirwana Putri / 21081010187

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
2024**

## Kata Pengantar

Puji syukur ke hadirat Allah SWT yang telah memberikan rahmat, nikmat, dan hidayah-Nya sehingga penulis dapat menyelesaikan proposal penelitian dengan judul "**Model Kombinasi Optimalisasi PCA dengan Teknik Adaptive Weighting untuk Deteksi Anomali pada Dataset Multivariat**". Proposal ini disusun sebagai salah satu syarat untuk melanjutkan proses akademik dalam program studi yang sedang dijalani.

Dalam penyusunan proposal ini, penulis mendapatkan banyak dukungan, motivasi, dan masukan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan rasa terima kasih kepada:

1. **Dr. Basuki Rahmat, S.Si., MT**, selaku dosen pembimbing yang telah meluangkan waktu untuk memberikan bimbingan, arahan, dan masukan yang sangat berharga selama proses penyusunan proposal ini.
2. **Universitas Pembangunan Nasional “Veteran” Jawa Timur**, sebagai tempat penulis menimba ilmu dan mendapatkan fasilitas untuk mendukung penelitian ini.
3. Orang tua dan keluarga tercinta, atas segala doa, dukungan, dan motivasi yang tiada henti diberikan kepada penulis.
4. Teman-teman seperjuangan di **Informatika**, yang telah menjadi tempat bertukar pikiran dan saling mendukung dalam proses akademik ini.

Penulis menyadari bahwa proposal ini masih memiliki banyak kekurangan dan jauh dari sempurna. Oleh karena itu, penulis dengan tangan terbuka menerima kritik dan saran yang membangun untuk penyempurnaan penelitian ini di masa mendatang. Semoga proposal ini dapat memberikan manfaat dan menjadi inspirasi bagi penelitian lain di bidang yang sama.

Akhir kata, penulis berharap semoga Allah SWT senantiasa memberikan berkah dan kemudahan dalam setiap langkah kita semua.

## Daftar isi

Kata Pengantar .....	2
Daftar Isi .....	3
Bab I   Pendahuluan.....	4
I.1   Latar Belakang .....	4
I.2   Rumusan Masalah.....	4
I.3   Tujuan Penelitian .....	5
I.4   Manfaat Penelitian .....	5
Bab II   Tinjauan Pustaka .....	6
II.1   Tinjauan Teoritis.....	6
II.2   Keunggulan dan Keterbatasan PCA.....	8
II.3   Ilustrasi Langkah-Langkah PCA.....	12
II.4   Perkembangan Adaptive Weighting .....	16
II.5   Penelitian Terkait .....	18
II.6   Research Gap .....	18
Bab III  Metodologi Penelitian.....	19
III.1  Desain Penelitian .....	19
III.2  Tahapan Penelitian.....	21
III.2.1 Pengumpulan Data .....	19
III.2.2 Preprocessing Data.....	21
III.2.3 Integrasi PCA dan Adaptive Weighting .....	19
III.2.4 Penerapan Model Deteksi Anomali .....	21
III.3  Evaluasi Model .....	24
III.3.1 Metode Evaluasi.....	25
III.4  Alat dan Software .....	25
III.5  Kerangka Kerja Penelitian .....	27
III.6  Batasan Penelitian.....	28

# **BAB I Pendahuluan**

## **I.1 Latar Belakang**

Pada era digital ini, data menjadi aset yang sangat berharga di berbagai sektor, termasuk kesehatan, keuangan, transportasi, dan industri. Volume data yang terus meningkat menimbulkan tantangan dalam analisis data multivariat, terutama ketika data memiliki dimensi tinggi (high-dimensional data). Data dengan dimensi tinggi seringkali mengandung redundansi informasi, di mana banyak fitur tidak relevan atau hanya memberikan sedikit kontribusi terhadap tujuan analisis [1]. Masalah ini dapat menghambat efisiensi komputasi, meningkatkan risiko overfitting pada model pembelajaran mesin, dan memperumit interpretasi hasil analisis.

Principal Component Analysis (PCA) adalah salah satu teknik yang sering digunakan untuk reduksi dimensi [2]. Metode ini bekerja dengan mentransformasikan data ke ruang baru, di mana sebagian besar varians data terkandung dalam beberapa komponen utama. Meskipun PCA sangat efektif dalam mengurangi dimensi, metode ini tidak mempertimbangkan relevansi spesifik dari fitur terhadap tujuan analisis tertentu. Dalam beberapa kasus, kehilangan fitur yang relevan dapat berdampak negatif pada performa model deteksi anomali.

Untuk mengatasi keterbatasan ini, penelitian ini mengusulkan integrasi PCA dengan teknik adaptive weighting. Teknik ini memungkinkan pemberian bobot yang lebih besar pada fitur yang dianggap penting, sehingga menghasilkan reduksi dimensi yang lebih efisien tanpa mengorbankan informasi kritis. Pendekatan ini menawarkan inovasi baru dalam analisis data multivariat dan memberikan solusi yang lebih adaptif untuk deteksi anomali di berbagai aplikasi dunia nyata.

## **I.2 Rumusan Masalah**

1. Bagaimana teknik adaptive weighting dapat meningkatkan efisiensi dan efektivitas PCA dalam reduksi dimensi dataset multivariat?
2. Apakah integrasi PCA dengan adaptive weighting dapat menghasilkan performa deteksi anomali yang lebih baik dibandingkan metode konvensional?

3. Bagaimana pengaruh parameter explained variance ratio dan bobot adaptif terhadap keseimbangan antara efisiensi komputasi dan akurasi deteksi anomali?

### **I.3 Tujuan Penelitian**

1. Mengembangkan model kombinasi PCA dengan teknik adaptive weighting untuk menghasilkan reduksi dimensi yang lebih relevan dan efisien.
2. Mengevaluasi performa model yang diusulkan dalam mendeteksi anomali pada berbagai jenis dataset multivariat.
3. Menentukan konfigurasi optimal parameter explained variance ratio dan bobot adaptif untuk mencapai hasil terbaik dalam analisis data.

### **I.4 Manfaat Penelitian**

1. Teoretis

Memberikan kontribusi terhadap pengembangan metode analisis data dengan mengintegrasikan pendekatan statistik (PCA) dan teknik adaptif (adaptive weighting).

2. Praktis

Menawarkan solusi praktis yang dapat diterapkan dalam berbagai aplikasi dunia nyata, seperti deteksi penipuan (fraud detection), analisis kesehatan, dan monitoring sistem industri.

3. Inovasi

Membuka peluang penelitian lanjutan untuk mengintegrasikan metode reduksi dimensi dengan teknik pembelajaran mesin dan algoritma adaptif lainnya.

## **BAB II Tinjauan Pustaka**

### **II.1 Tinjauan Teoritis**

Principal Component Analysis (PCA) adalah sebuah metode statistik yang digunakan untuk mereduksi dimensi dataset dengan mentransformasikan data ke dalam ruang koordinat baru, yang disebut komponen utama. Dengan menggunakan PCA, kita dapat menyederhanakan data berukuran besar tanpa kehilangan informasi yang signifikan. Metode ini sangat populer dalam analisis data karena kemampuannya untuk mengidentifikasi pola tersembunyi dalam data dan membuatnya lebih mudah untuk dianalisis, dipahami, dan divisualisasikan.

PCA berfungsi untuk mengidentifikasi arah variasi terbesar dalam dataset dan mereduksi dimensi data dengan memproyeksikannya ke dalam ruang baru. Ruang baru ini dibentuk oleh komponen utama, yang merupakan kombinasi linier dari fitur asli dalam dataset. Proses ini penting ketika kita memiliki data yang sangat besar, yang mungkin memiliki banyak fitur dan variabel yang saling berkorelasi. Dengan menggunakan PCA, kita dapat mengurangi jumlah fitur yang diperlukan untuk menggambarkan data, sambil mempertahankan sebagian besar informasi yang ada.

PCA melibatkan beberapa langkah utama yang harus diikuti untuk mendapatkan hasil yang optimal:

#### **1. Standarisasi Data**

Salah satu hal pertama yang perlu dilakukan sebelum menjalankan PCA adalah standarisasi data. Hal ini dilakukan untuk memastikan bahwa semua fitur memiliki skala yang sama dan tidak ada fitur yang mendominasi analisis karena perbedaan skala. Sebagai contoh, dalam dataset yang mencakup dua fitur, seperti berat badan (dalam kilogram) dan tinggi badan (dalam sentimeter), fitur berat badan akan memiliki skala yang jauh lebih besar daripada tinggi badan. Jika kita tidak menstandarisasi data, fitur dengan skala yang lebih besar (misalnya, berat badan) akan lebih mempengaruhi hasil PCA.

Proses standarisasi mengubah setiap fitur agar memiliki rata-rata nol dan varians satu. Ini memastikan bahwa semua fitur memiliki kontribusi yang setara terhadap analisis,

tanpa dipengaruhi oleh perbedaan skala.

## 2. Perhitungan Matriks Kovarian

Setelah data distandarisasi, langkah berikutnya adalah menghitung matriks kovarian. Matriks kovarian menunjukkan hubungan linier antara setiap pasangan fitur dalam dataset. Setiap elemen dalam matriks ini mewakili kovarian antara dua fitur, yang mencerminkan sejauh mana perubahan satu fitur berhubungan dengan perubahan fitur lainnya.

Jika dua fitur memiliki kovarian yang tinggi, artinya keduanya bergerak dalam arah yang sama. Sebaliknya, jika kovariannya rendah atau negatif, kedua fitur tersebut tidak berkorelasi erat. Matriks kovarian ini sangat penting karena PCA berfokus pada mencari komponen utama yang menggambarkan variasi terbesar dalam data, yang berkaitan erat dengan hubungan antar fitur dalam dataset.

## 3. Dekomposisi Eigen

Langkah berikutnya dalam PCA adalah dekomposisi eigen dari matriks kovarian. Dekomposisi ini menghasilkan dua hal penting:

- **Eigenvalue** : Menunjukkan seberapa besar varians yang dijelaskan oleh setiap komponen utama. Komponen utama dengan eigenvalue yang lebih besar menjelaskan lebih banyak informasi atau variasi dalam data.
- **Eigenvector** : Menentukan arah atau orientasi dari setiap komponen utama dalam ruang fitur asli. Eigenvector ini merupakan kombinasi linier dari fitur-fitur yang ada dalam dataset dan mengindikasikan bagaimana data diproyeksikan ke dalam ruang baru.

Eigenvalue dan eigenvector ini digunakan untuk menentukan seberapa penting setiap komponen utama dalam menjelaskan variasi data. Eigenvalue yang lebih besar menunjukkan komponen utama yang lebih penting.

## 4. Pemilihan Komponen Utama

Setelah dekomposisi eigen, langkah berikutnya adalah memilih komponen utama

yang paling relevan untuk data. Salah satu cara untuk menentukan komponen utama yang penting adalah dengan melihat explained variance ratio. Explained variance ratio mengukur seberapa banyak varians dalam data yang dijelaskan oleh setiap komponen utama.

Explained variance ratio dihitung dengan membagi eigenvalue setiap komponen dengan jumlah total eigenvalue. Komponen utama yang memiliki explained variance ratio yang lebih tinggi akan dipilih karena mereka menyumbang lebih banyak informasi tentang dataset asli. Biasanya, hanya beberapa komponen utama yang dipilih, yang cukup untuk menjelaskan sebagian besar varians dalam data. Dengan memilih komponen utama yang relevan, kita dapat mereduksi dimensi data secara signifikan tanpa kehilangan banyak informasi penting.

## 5. Proyeksi Data

Setelah memilih komponen utama, langkah terakhir adalah memproyeksikan data ke dalam ruang baru yang dibentuk oleh eigenvector yang terpilih. Proyeksi ini menghasilkan dataset dengan dimensi lebih rendah yang mempertahankan sebagian besar informasi yang ada dalam data asli.

Misalnya, jika dataset awal memiliki 10 fitur, PCA dapat mengurangi dimensi data menjadi 2 atau 3 dimensi, tergantung pada komponen utama yang dipilih. Meskipun dimensi data dikurangi, informasi yang paling relevan tetap terjaga, dan dataset baru ini lebih mudah dianalisis dan divisualisasikan.

## II.2 Keunggulan dan Keterbatasan PCA

Berikut adalah penjelasan yang lebih mendalam mengenai keunggulan dan keterbatasan Principal Component Analysis (PCA):

### Keunggulan PCA

#### 1. Mengurangi Redundansi Data

Salah satu keunggulan utama PCA adalah kemampuannya untuk mengurangi



redundansi data. Banyak dataset memiliki variabel yang saling berkorelasi (redundansi). PCA membantu mengidentifikasi dan mengurangi korelasi antar variabel dengan mentransformasikan data ke dalam komponen utama (principal components) yang saling ortogonal satu sama lain. Hal ini memungkinkan kita untuk mengurangi jumlah variabel tanpa kehilangan informasi yang signifikan. Contohnya dalam dataset gambar, banyak fitur (seperti piksel) mungkin sangat berkorelasi. PCA mengubah gambar ini menjadi sejumlah komponen utama yang lebih terisolasi, mengurangi redundansi dan meningkatkan efisiensi analisis.

## 2. Meningkatkan Efisiensi Komputasi

Dengan mengurangi dimensi data (dimensionality reduction), PCA dapat meningkatkan efisiensi komputasi, terutama ketika dataset memiliki banyak fitur. Model machine learning yang dilatih dengan data berdimensi rendah akan lebih cepat dalam hal waktu pelatihan dan prediksi, serta lebih efisien dalam penggunaan sumber daya komputasi. Contohnya dalam analisis data besar seperti genomik atau analisis data sensor, PCA memungkinkan untuk mengurangi jumlah fitur yang harus diproses, sehingga mempercepat algoritma pembelajaran mesin dan mengurangi konsumsi memori.

## 3. Memvisualisasikan Dataset yang Kompleks

PCA juga sering digunakan untuk memvisualisasikan data berdimensi tinggi dalam ruang dua atau tiga dimensi. Dengan mereduksi data yang awalnya memiliki banyak dimensi (misalnya ratusan atau ribuan fitur) menjadi dua atau tiga komponen utama, PCA membuatnya lebih mudah untuk memahami pola dan hubungan dalam data. Contohnya dalam analisis citra atau pengelompokan data, PCA dapat digunakan untuk mengubah data berdimensi tinggi menjadi visualisasi dua dimensi yang mudah dipahami, memungkinkan identifikasi pola atau kluster dalam data.

### Keterbatasan PCA

#### 1. Asumsi Linearitas

PCA mengasumsikan bahwa hubungan antara variabel-variabel dalam data adalah linear. Ini berarti bahwa PCA hanya dapat menangkap hubungan linier antara fitur. Jika dataset memiliki hubungan non-linear yang kompleks, PCA mungkin tidak dapat

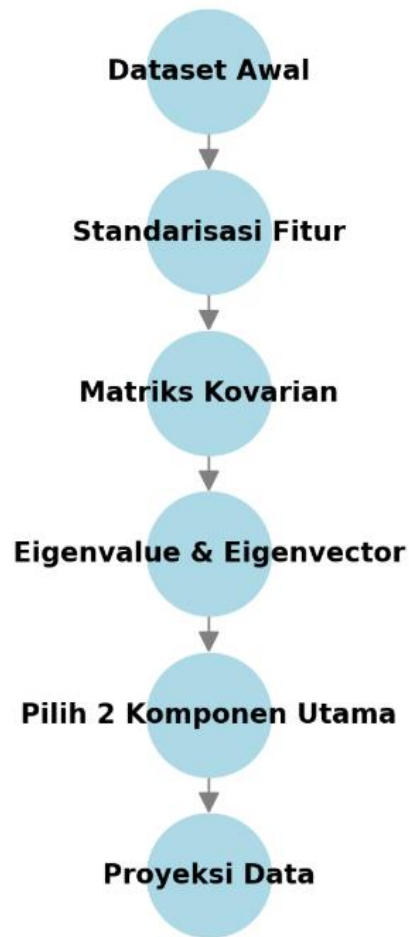
menggambarkan pola yang sebenarnya ada dalam data. Contohnya dalam pengolahan citra atau data sensor yang sangat tidak linier (misalnya pola dalam gambar dengan perubahan bentuk atau tekstur yang kompleks), PCA mungkin tidak dapat menangkap informasi yang penting karena hubungan antar fitur tidak linear.

## 2. Tidak Mampu Menangani Data Non-Linear Secara Langsung

PCA memiliki keterbatasan dalam menangani struktur data non-linear secara langsung. Meskipun PCA sangat efektif dalam mengurangi dimensi data yang bersifat linier, ia tidak dapat secara langsung menangani data yang memiliki hubungan non-linier tanpa modifikasi atau teknik tambahan. Contohnya pada data yang berasal dari manifold non-linear, seperti pengenalan wajah dengan variasi pencahayaan atau sudut pandang, PCA cenderung gagal dalam menangkap informasi yang relevan karena ia hanya berfokus pada hubungan linier antar fitur. Untuk menangani hal ini, teknik seperti Kernel PCA atau t-SNE (t-distributed Stochastic Neighbor Embedding) sering digunakan untuk menangkap hubungan non-linier.

PCA adalah alat yang sangat berguna dalam analisis data, terutama untuk mengurangi dimensi, menghilangkan redundansi, dan mempercepat proses komputasi. Namun, keterbatasannya dalam menangani data dengan hubungan non-linear dan asumsi linearitas menjadikannya kurang cocok untuk beberapa jenis dataset, terutama yang memiliki struktur kompleks atau non-linear. Untuk data yang lebih kompleks, pendekatan lain seperti Kernel PCA atau Deep Learning mungkin lebih efektif dalam menangani data yang memiliki pola non-linear.

### II.3 Ilustrasi Langkah-Langkah PCA



Gambar 2.1 Diagram alur PCA

Berikut adalah penjelasan yang lebih detail tentang langkah-langkah dalam Principal Component Analysis (PCA), berdasarkan tahapannya:

#### 1. Dataset Awal diambil dari Berbagai Sumber dengan Fitur yang Memiliki Skala Berbeda

Dataset yang digunakan dalam PCA biasanya berasal dari berbagai sumber atau eksperimen yang mengumpulkan data dengan berbagai fitur atau variabel. Fitur-fitur ini sering kali memiliki skala yang sangat berbeda satu sama lain. Sebagai contoh, dalam dataset yang mencakup informasi tentang produk, kita mungkin memiliki fitur-fitur seperti:

- Harga Produk: Dapat memiliki rentang nilai antara 100 hingga 1.000.000 (atau lebih).

- Jumlah Produk Terjual: Dapat berkisar antara 0 hingga 500.000 unit.
- Rating Produk: Biasanya terletak dalam rentang 1 hingga 5.

Fitur-fitur ini memiliki skala yang sangat berbeda, yang dapat membuat analisis lebih sulit dan menyebabkan beberapa fitur mendominasi hasil analisis. Penting untuk memastikan bahwa tidak ada fitur yang mendominasi karena skala yang lebih besar.

## 2. Dilakukan Standarisasi Sehingga Setiap Fitur Memiliki Rata-Rata Nol dan Varians Satu

Langkah pertama yang penting sebelum melakukan PCA adalah standarisasi data. Proses ini dilakukan untuk memastikan bahwa setiap fitur berada dalam skala yang sama, sehingga tidak ada fitur yang lebih dominan hanya karena memiliki nilai yang lebih besar.

Proses standarisasi dilakukan dengan rumus berikut untuk setiap fitur  $x_i$  dalam dataset:

$$z_i = \frac{x_i - \mu}{\sigma}$$

Di mana:

- $x_i$  adalah nilai dari fitur pada sampel ke- $i$ ,
- $\mu$  adalah rata-rata (mean) dari fitur tersebut,
- $\sigma$  adalah standar deviasi dari fitur tersebut,
- $z_i$  adalah nilai standar (terstandarisasi) dari fitur.

Proses ini mengubah setiap fitur menjadi distribusi yang memiliki rata-rata nol dan varians satu, sehingga setiap fitur akan memiliki kontribusi yang sama terhadap analisis selanjutnya.

## 3. Matriks Kovarian Dihitung, Misalnya untuk Dataset dengan Tiga Fitur, Menghasilkan Matriks 3x3

Setelah standarisasi, langkah berikutnya adalah menghitung matriks kovarian dari data yang telah distandarisasi. Matriks kovarian digunakan untuk menggambarkan hubungan linier antar fitur dalam dataset. Elemen-elemen dalam matriks ini menunjukkan seberapa besar dua fitur berubah bersama (kovariansi).

Jika dataset memiliki  $n$  fitur, maka matriks kovarian akan menjadi matriks  $n \times n$ . Sebagai contoh, jika dataset memiliki tiga fitur, maka matriks kovarian akan menjadi matriks  $3 \times 3$ , dengan elemen-elemen yang dihitung sebagai berikut:

- Elemen diagonal  $\text{cov}(x_i, x_i)$  adalah varians dari fitur  $x_i$ ,
- Elemen non-diagonal  $\text{cov}(x_i, x_j)$  adalah kovarians antara fitur  $x_i$  dan  $x_j$ .

Matriks kovarian untuk dataset tiga fitur (misalnya, Harga Produk, Jumlah Produk Terjual, dan Rating Produk) akan terlihat seperti ini:

$$\text{Covariance Matrix} = \begin{bmatrix} \text{COV}(x_1, x_1) & \text{COV}(x_1, x_2) & \text{COV}(x_1, x_3) \\ \text{COV}(x_2, x_1) & \text{COV}(x_2, x_2) & \text{COV}(x_2, x_3) \\ \text{COV}(x_3, x_1) & \text{COV}(x_3, x_2) & \text{COV}(x_3, x_3) \end{bmatrix}$$

Matriks ini mencerminkan sejauh mana perubahan satu fitur terkait dengan perubahan fitur lainnya. Jika dua fitur memiliki kovarians tinggi, artinya mereka berkorelasi kuat.

#### 4. Eigenvalue dan Eigenvector Diekstraksi dari Matriks Kovarian

Setelah matriks kovarian dihitung, langkah selanjutnya adalah dekomposisi eigen. Dekomposisi ini bertujuan untuk menguraikan matriks kovarian menjadi eigenvalue dan eigenvector.

- Eigenvalue ( $\lambda$ )

Mengukur seberapa besar varians yang dijelaskan oleh setiap komponen utama. Eigenvalue yang lebih besar menunjukkan komponen utama yang lebih penting, karena ia menjelaskan lebih banyak variasi dalam data.

- Eigenvector ( $v$ )

Mengindikasikan arah dari komponen utama. Eigenvector ini adalah kombinasi linier dari fitur-fitur dalam data, yang menunjukkan arah di mana data tersebar.

Proses ini menghasilkan satu set eigenvalue dan eigenvector yang masing-masing berpasangan. Eigenvalue yang lebih besar mengindikasikan bahwa komponen utama yang terkait dengan eigenvector tersebut membawa lebih banyak informasi atau varians dari data.

#### 5. Dua Komponen Utama Pertama dengan Eigenvalue Terbesar Dipilih

Setelah mendapatkan eigenvalue dan eigenvector, langkah berikutnya adalah memilih komponen utama yang paling penting. Komponen utama yang dipilih berdasarkan eigenvalue terbesar, karena komponen-komponen ini menjelaskan varians terbesar dalam data.

- Misalnya, jika kita memiliki tiga komponen utama, kita akan memilih dua komponen pertama yang memiliki eigenvalue terbesar. Komponen utama ini menjelaskan sebagian besar variasi dalam dataset, sehingga tetap menjaga informasi yang paling penting.
- Explained Variance Ratio digunakan untuk menentukan seberapa banyak varians yang dijelaskan oleh masing-masing komponen utama. Biasanya, kita memilih komponen utama yang cumulatively menjelaskan lebih dari 80% hingga 90% dari varians data.

#### 6. Data Diproyeksikan ke Ruang Dua Dimensi yang Dibentuk oleh Dua Komponen Utama Ini

Langkah terakhir dalam PCA adalah proyeksi data ke dalam ruang baru yang dibentuk oleh dua komponen utama yang telah dipilih. Data yang sebelumnya memiliki banyak dimensi akan diproyeksikan ke dalam dua dimensi berdasarkan dua eigenvector terpilih.

Proyeksi ini dapat dihitung dengan mengalikan data yang telah distandarisi dengan matriks yang dibentuk dari dua eigenvector yang terpilih. Dengan kata lain, kita mengalikan data yang sudah distandarisi dengan matriks yang berisi dua eigenvector untuk mendapatkan data dalam dimensi yang lebih rendah.

Hasilnya adalah dataset dengan dua dimensi yang masih mempertahankan sebagian besar informasi dan varians dari dataset asli. Visualisasi data ini dalam dua dimensi sangat berguna untuk analisis lebih lanjut, seperti pengelompokan (clustering) atau visualisasi pola dalam data.

Proses PCA membantu mereduksi dimensi dataset dengan cara yang sangat efisien, memungkinkan kita untuk mempertahankan informasi penting sambil mengurangi kompleksitas

data. Dengan mengikuti langkah-langkah ini, kita dapat memperoleh pemahaman yang lebih baik tentang struktur data dan menyederhanakan analisis selanjutnya.

## **II.4 Perkembangan Adaptive Weighting**

Perkembangan Adaptive Weighting merujuk pada teknik yang digunakan untuk menentukan bobot yang dinamis dan adaptif dalam pemrosesan data atau model. Teknik ini dapat digunakan dalam berbagai aplikasi analisis dan machine learning, di mana bobot fitur, data, atau komponen dihitung berdasarkan kepentingan atau kontribusinya terhadap hasil akhir. Adaptive weighting mengutamakan pengaturan bobot yang lebih cerdas, bukan hanya bergantung pada nilai tetap, sehingga dapat lebih responsif terhadap perubahan dalam data atau konteks.

Berikut adalah penjelasan lebih mendalam tentang perkembangan adaptive weighting dalam beberapa bidang utama:

### **1. Image Recognition**

Dalam pengenalan gambar (image recognition), adaptive weighting digunakan untuk meningkatkan kinerja model dengan memberi bobot lebih pada fitur yang lebih penting dalam mendeteksi objek tertentu. Ini membantu model untuk lebih fokus pada elemen-elemen penting dalam gambar, seperti tepi atau pola yang menonjol, daripada elemen-elemen yang kurang relevan. Teknik ini memungkinkan jaringan saraf (neural networks) untuk belajar lebih efisien dan mengenali objek dengan lebih baik.

Contoh penerapan:

- Convolutional Neural Networks (CNNs) sering menggunakan metode adaptive weighting untuk menekankan area gambar yang lebih informatif (misalnya, wajah pada gambar wajah atau kendaraan pada gambar lalu lintas).
- Dalam transfer learning, ketika model yang sudah dilatih pada satu set data digunakan untuk tugas lain, adaptive weighting bisa menyesuaikan bobot filter dalam model agar sesuai dengan data baru yang lebih spesifik.

### **2. Text Classification**

Pada klasifikasi teks, adaptive weighting berfungsi untuk memberi bobot lebih pada kata-kata atau fitur tertentu yang memberikan kontribusi lebih besar terhadap hasil klasifikasi. Misalnya, dalam analisis sentimen atau klasifikasi email, beberapa kata kunci mungkin lebih menentukan dalam menentukan kategori teks, seperti kata "happy" untuk sentimen positif atau "urgent" untuk email penting.

Teknik yang digunakan:

- Information Gain adalah salah satu metode yang sering digunakan dalam adaptive weighting untuk menghitung kontribusi relatif setiap fitur dalam klasifikasi. Fitur dengan gain tertinggi akan diberi bobot lebih besar karena mereka lebih informatif dalam menentukan kelas teks.
- TF-IDF (Term Frequency-Inverse Document Frequency) adalah contoh lainnya, di mana kata-kata yang lebih jarang muncul di seluruh korpus teks namun sering muncul dalam sebuah dokumen diberi bobot lebih besar, karena mereka memberikan informasi yang lebih relevan untuk klasifikasi.

### 3. Analitik Bisnis

Dalam analitik bisnis, adaptive weighting digunakan untuk meningkatkan akurasi model prediktif dengan memberi bobot lebih pada fitur-fitur yang lebih relevan dengan tujuan analisis. Misalnya, dalam pemasaran atau analisis risiko, adaptive weighting membantu untuk memahami faktor-faktor yang lebih penting dalam menentukan hasil (seperti penjualan atau potensi risiko).

Contoh penerapan:

- Analisis churn: Dalam konteks analisis churn pelanggan, bobot lebih besar dapat diberikan pada faktor-faktor seperti penggunaan produk, interaksi dengan layanan pelanggan, atau kepuasan yang diukur melalui survei.
- Analisis risiko finansial: Dalam model prediksi kredit, fitur seperti pendapatan, skor kredit, dan riwayat pembayaran dapat diberi bobot lebih besar karena lebih relevan dalam memprediksi kelayakan kredit.

Secara keseluruhan, penerapan adaptive weighting yang efisien memungkinkan model



dalam berbagai domain untuk lebih fokus pada fitur yang paling relevan, meningkatkan akurasi, dan mengurangi kompleksitas komputasi dengan menurunkan pengaruh fitur yang tidak begitu penting. Teknik ini menjadi semakin penting seiring dengan perkembangan teknologi machine learning dan AI yang semakin canggih.

## **II.5 Penelitian Terkait**

Optimalisasi PCA telah menjadi fokus utama dalam berbagai studi. Sebagian besar penelitian menggunakan explained variance ratio sebagai parameter untuk menentukan jumlah komponen yang dipilih. Contohnya, penelitian sebelumnya menunjukkan bahwa memilih komponen dengan explained variance ratio kumulatif di atas 90% dapat mempertahankan informasi penting dari dataset. Namun, penelitian ini sering mengabaikan peran adaptasi bobot fitur untuk meningkatkan kualitas reduksi dimensi.

Studi lain telah menggabungkan PCA dengan metode ekstraksi fitur, seperti Histogram of Oriented Gradients (HOG) dan Local Binary Patterns (LBP), untuk meningkatkan performa model analitik. Kombinasi ini terbukti efektif dalam aplikasi pengenalan pola dan visi komputer. Dalam konteks deteksi anomali, algoritma seperti Isolation Forest, DBSCAN, dan Autoencoders telah banyak digunakan. Namun, sebagian besar penelitian ini tidak mengeksplorasi dampak reduksi dimensi terhadap kinerja deteksi anomali.

## **II.6 Research Gap**

Meskipun PCA telah lama digunakan sebagai teknik reduksi dimensi, belum ada pendekatan eksplisit yang mengintegrasikan PCA dengan adaptive weighting untuk meningkatkan relevansi fitur dalam konteks deteksi anomali. Penelitian sebelumnya cenderung memisahkan proses reduksi dimensi dan pembobotan fitur, sehingga peluang untuk menciptakan representasi data yang lebih informatif belum dimanfaatkan secara maksimal. Selain itu, dampak dari penggunaan dataset berdimensi rendah hasil PCA terhadap algoritma deteksi anomali belum diteliti secara mendalam.

Penelitian ini bertujuan untuk mengisi celah tersebut dengan mengusulkan integrasi PCA dan adaptive weighting sebagai pendekatan baru yang lebih adaptif dan relevan terhadap karakteristik dataset multivariat. Dengan pendekatan ini, diharapkan deteksi anomali dapat

dilakukan dengan akurasi lebih tinggi, efisiensi komputasi yang lebih baik, serta kemampuan untuk menangani dataset yang kompleks secara lebih efektif.

## **BAB III Metodologi Penelitian**

### **III.1 Desain Penelitian**

Penelitian ini menggunakan metode eksperimen sebagai pendekatan utama untuk mengembangkan dan menguji model kombinasi Principal Component Analysis (PCA) dengan teknik adaptive weighting. Metode eksperimen dipilih karena memberikan kendali penuh terhadap proses pengujian dan pengembangan model, yang mencakup analisis sistematis atas setiap tahap dan variabel yang mempengaruhi keberhasilan deteksi anomali pada dataset multivariat. Desain penelitian dirancang untuk mengevaluasi kemampuan model secara terstruktur, dari mulai tahap preprocessing data hingga pengukuran kinerja model pada berbagai konfigurasi dataset.

Tahap awal dari desain penelitian ini adalah merancang eksperimen yang mampu menangkap kompleksitas dataset multivariat, terutama yang memiliki dimensi tinggi (high-dimensional data). Dataset dengan dimensi tinggi seringkali mengandung fitur-fitur yang redundan atau tidak relevan, sehingga reduksi dimensi menjadi salah satu langkah penting dalam eksperimen ini. PCA digunakan sebagai metode utama untuk mereduksi dimensi dataset [3]. Namun, dalam penelitian ini, PCA tidak diterapkan secara tradisional, melainkan dioptimalkan dengan teknik adaptive weighting. Integrasi ini dirancang untuk memastikan bahwa setiap fitur dalam dataset mendapatkan perhatian sesuai dengan tingkat relevansinya terhadap deteksi anomali, sehingga hasil reduksi dimensi tetap informatif dan efisien.

Desain eksperimen juga mempertimbangkan bahwa dataset multivariat seringkali memiliki tantangan berupa distribusi yang kompleks, tingkat korelasi antar fitur yang tinggi, serta keberadaan anomali yang sulit terdeteksi. Oleh karena itu, proses eksperimen dirancang untuk mengevaluasi apakah model yang diusulkan mampu mengatasi tantangan ini dengan menghasilkan deteksi anomali yang akurat dan konsisten. Dalam desain ini, berbagai skenario dataset digunakan, termasuk dataset dengan distribusi normal, non-normal, dan yang mengandung proporsi anomali berbeda-beda.

Tahapan dalam desain eksperimen mencakup proses berikut:

1. Perencanaan Eksperimen

Peneliti menentukan parameter utama yang akan diuji dalam eksperimen, seperti explained variance ratio pada PCA dan bobot adaptif yang digunakan dalam teknik adaptive weighting. Parameter-parameter ini dipilih berdasarkan signifikansinya terhadap performa model, sehingga setiap perubahan pada parameter dapat menunjukkan pengaruh hasil.

## 2. Preprocessing Data

Proses ini mencakup penanganan data mentah untuk memastikan bahwa dataset siap digunakan dalam eksperimen. Semua data distandarisasi agar setiap fitur memiliki skala yang sama, sehingga tidak ada fitur yang mendominasi analisis karena skala yang lebih besar. Selain itu, jika terdapat data yang hilang, langkah-langkah seperti imputasi, dilakukan untuk mengisi kekosongan ini. Standarisasi dan imputasi dilakukan untuk dataset.

## 3. Penerapan PCA dengan Teknik Adaptive Weighting

PCA diterapkan pada dataset untuk mereduksi dimensi dengan memilih komponen utama yang menjelaskan sebagian besar varians data. Dalam penelitian ini, PCA tidak hanya diterapkan sebagai alat reduksi dimensi, tetapi juga diintegrasikan dengan adaptive weighting. Adaptive weighting memberikan bobot lebih besar pada fitur yang memiliki relevansi tinggi terhadap deteksi anomali, memastikan bahwa hasil reduksi dimensi adaptif.

## 4. Evaluasi Model

Setelah model diterapkan, performanya dievaluasi pada dataset uji yang belum pernah dilihat oleh model sebelumnya. Evaluasi dilakukan dengan mengukur metrik akurasi, seperti precision, recall, dan F1-score, untuk memastikan bahwa model mampu mendeteksi anomali secara konsisten. Selain itu, efisiensi komputasi model diukur untuk mengevaluasi apakah integrasi PCA dan adaptive weighting mampu meningkatkan kecepatan dan mengurangi kebutuhan sumber daya dibandingkan metode tradisional.

Desain penelitian ini juga mencakup eksperimen parameter untuk mengidentifikasi konfigurasi optimal dari explained variance ratio dan bobot adaptif. Eksperimen parameter ini dirancang untuk memastikan bahwa model yang dihasilkan tidak hanya akurat, tetapi juga efisien dalam konteks komputasi. Setiap tahap dalam desain penelitian ini dilaksanakan dengan hati-hati untuk memastikan bahwa hasil yang diperoleh valid dan dapat diterapkan dalam berbagai konteks dunia nyata.

Dengan desain penelitian yang sistematis ini, penelitian bertujuan untuk menguji keunggulan model yang diusulkan dan memberikan kontribusi signifikan dalam pengembangan metode deteksi anomali berbasis PCA dan adaptive weighting [4]. Desain ini mencakup semua aspek penting dalam pengembangan model, mulai dari preprocessing data hingga evaluasi hasil, sehingga memberikan gambaran komprehensif tentang efektivitas pendekatan yang digunakan.

## **III.2 Tahapan Penelitian**

Tahapan penelitian dalam proyek ini dirancang untuk memastikan bahwa setiap langkah proses memiliki tujuan yang jelas dan mendukung keberhasilan model deteksi anomali. Tahapan ini dimulai dari pengumpulan data, diikuti oleh preprocessing data, penerapan PCA dengan adaptive weighting, hingga evaluasi performa model. Berikut adalah penjelasan detail untuk setiap tahap:

### **III.2.1 Pengumpulan Data**

Tahap pengumpulan data merupakan langkah awal dalam penelitian ini. Dataset multivariat dikumpulkan dari sumber terpercaya seperti repositori publik (misalnya UCI Machine Learning Repository) yang menyediakan data dengan variasi fitur, kompleksitas, dan ground truth yang sesuai untuk evaluasi deteksi anomali. Pemilihan dataset dilakukan berdasarkan beberapa kriteria, termasuk:

1. Kompleksitas Dimensi

Dataset harus memiliki jumlah fitur yang cukup tinggi untuk menguji efektivitas PCA sebagai metode reduksi dimensi.

## 2. Variasi Jenis Fitur

Kombinasi fitur numerik dan kategorikal memberikan tantangan tambahan dalam preprocessing.

## 3. Ground Truth

Keberadaan label atau informasi anomali dalam dataset digunakan sebagai referensi untuk evaluasi model.

### **III.2.2 Preprocessing Data**

Preprocessing adalah tahap penting untuk memastikan bahwa data mentah dapat digunakan secara efektif dalam model. Proses ini mencakup beberapa langkah berikut:

#### 1. Standarisasi Data

Setiap fitur dalam dataset distandarisasi agar memiliki rata-rata 0 dan standar deviasi 1 [5]. Standarisasi dilakukan untuk menghilangkan bias akibat skala fitur yang berbeda. Formula yang digunakan di mana adalah mean fitur dan adalah standar deviasi fitur.

#### 2. Penanganan Missing Values

Data yang hilang diimputasi menggunakan metode mean untuk fitur numerik dan modus untuk fitur kategorikal. Jika proporsi data yang hilang terlalu tinggi (misalnya lebih dari 30%), fitur tersebut dapat dihapus untuk menghindari distorsi dalam analisis.

#### 3. Reduksi Dimensi Awal

PCA diterapkan untuk mereduksi dimensi dataset dengan memilih komponen utama berdasarkan explained variance ratio di atas ambang tertentu (misalnya 90%). Tahap ini bertujuan untuk mengidentifikasi pola dasar dalam dataset sebelum integrasi dengan adaptive weighting.

### **III.2.3 Integrasi PCA dan Adaptive Weighting**

Setelah preprocessing selesai, tahap ini melibatkan penerapan PCA yang dioptimalkan dengan teknik adaptive weighting:

1. Penentuan Bobot

Bobot adaptif dihitung untuk setiap fitur berdasarkan relevansinya terhadap variabel target atau pola anomali. Korelasi Pearson digunakan untuk data numerik, sementara metode chi-square digunakan untuk data kategorikal. Formula bobot adaptif di mana adalah relevansi fitur.

2. Transformasi Data

PCA diterapkan kembali pada dataset yang telah diberikan bobot adaptif untuk menghasilkan representasi data yang lebih relevan dan informatif. Proses ini bertujuan untuk memastikan bahwa informasi penting dalam dataset tetap dipertahankan meskipun dimensinya telah direduksi.

### **III.2.4 Penerapan Model Deteksi Anomali**

Dataset hasil transformasi digunakan sebagai input untuk algoritma deteksi anomali, termasuk Isolation Forest, DBSCAN, dan Autoencoders. Isolation Forest bekerja dengan cara membangun pohon keputusan untuk mengisolasi data, di mana parameter utama seperti jumlah estimators dan ukuran sub-sample diatur untuk mengoptimalkan kinerja. Algoritma DBSCAN mendeteksi anomali berdasarkan kepadatan data, dengan parameter epsilon menentukan jarak maksimum antar titik untuk membentuk cluster, dan minPts menentukan jumlah minimum titik dalam cluster. Sementara itu, Autoencoders menggunakan jaringan saraf dengan lapisan tersembunyi untuk merekonstruksi data; error rekonstruksi yang tinggi menunjukkan potensi anomali. Parameter seperti ukuran lapisan tersembunyi, fungsi aktivasi, dan jumlah epochs diatur untuk memastikan model dapat belajar dengan baik dari pola data normal.

### **III.3 Evaluasi Model**

Evaluasi model merupakan tahap penting dalam penelitian ini untuk memastikan bahwa pendekatan yang diusulkan, yaitu kombinasi PCA dengan teknik adaptive weighting, dapat

memberikan hasil yang akurat dan efisien dalam mendeteksi anomali. Metode evaluasi yang digunakan mencakup pengukuran performa model dengan berbagai metrik dan pengujian efisiensi komputasi pada dataset dengan konfigurasi berbeda. Selain itu, eksperimen parameter dilakukan untuk memahami dampak variabel seperti explained variance ratio dan bobot adaptif terhadap performa keseluruhan model.

### **III.3.1 Metode Evaluasi**

Performa model dievaluasi menggunakan beberapa metrik utama, yaitu precision, recall, dan F1-score. Precision digunakan untuk mengukur sejauh mana prediksi anomali yang dibuat oleh model benar-benar merupakan anomali, sedangkan recall mengukur sejauh mana model mampu mendeteksi semua anomali yang ada dalam dataset. F1-score, yang merupakan rata-rata harmonis dari precision dan recall, digunakan sebagai metrik utama untuk menyeimbangkan keduanya, terutama pada dataset yang memiliki distribusi kelas yang tidak seimbang.

Selain itu, efisiensi komputasi diukur berdasarkan waktu eksekusi model dalam memproses dataset dengan berbagai tingkat reduksi dimensi. Pengukuran waktu eksekusi dilakukan untuk setiap tahap proses, mulai dari preprocessing data, penerapan PCA dengan adaptive weighting, hingga penerapan algoritma deteksi anomali seperti Isolation Forest, DBSCAN, dan Autoencoders. Evaluasi efisiensi ini penting untuk memastikan bahwa pendekatan yang diusulkan tidak hanya menghasilkan hasil yang akurat, tetapi juga dapat diimplementasikan secara praktis dalam skenario dunia nyata yang memerlukan kecepatan dan skalabilitas.

Eksperimen parameter dilakukan untuk mengevaluasi pengaruh explained variance ratio dan bobot adaptif terhadap performa model [6]. Explained variance ratio merupakan parameter utama dalam PCA yang menentukan jumlah komponen utama yang dipilih berdasarkan persentase varians data yang ingin dipertahankan. Dalam eksperimen ini, nilai explained variance ratio divariasikan dari 80% hingga 99% untuk mengamati dampaknya terhadap akurasi dan efisiensi model. Sementara itu, bobot adaptif dihitung berdasarkan relevansi fitur terhadap variabel target atau pola anomali, dan bobot ini divariasikan untuk memahami sejauh mana sensitivitas model terhadap perubahan bobot tersebut.

Eksperimen dilakukan pada beberapa dataset dengan karakteristik yang berbeda, seperti



jumlah fitur, tingkat sparsitas data, dan proporsi anomali. Hal ini dilakukan untuk mengevaluasi generalisasi model dalam berbagai skenario. Dataset yang digunakan mencakup dataset dengan distribusi normal, non-normal, dan dataset yang memiliki hubungan non-linier antar fitur. Pengujian dilakukan secara terpisah untuk setiap jenis dataset untuk mengidentifikasi keunggulan dan batasan pendekatan yang diusulkan.

Selama evaluasi, analisis mendalam dilakukan untuk memahami pola kesalahan yang terjadi, seperti false positive dan false negative dalam deteksi anomali. False positive terjadi ketika model secara salah mengidentifikasi data normal sebagai anomali, sedangkan false negative terjadi ketika model gagal mendeteksi data yang sebenarnya merupakan anomali. Analisis ini bertujuan untuk meningkatkan performa model dengan memperbaiki parameter atau menyesuaikan bobot adaptif dalam iterasi berikutnya.

Hasil dari evaluasi ini diharapkan dapat memberikan gambaran yang jelas tentang kelebihan dan kekurangan model, serta validasi terhadap klaim bahwa kombinasi PCA dengan teknik adaptive weighting dapat meningkatkan relevansi dan efisiensi dalam reduksi dimensi dan deteksi anomali [8].

### **III.4 Alat dan Software**

Penelitian ini menggunakan Python sebagai bahasa pemrograman utama karena fleksibilitas dan kelengkapan ekosistemnya dalam menangani berbagai kebutuhan analisis data. Python telah menjadi standar dalam penelitian berbasis data karena memiliki berbagai pustaka (library) yang mendukung pemrosesan data, analisis statistik, pembelajaran mesin, serta visualisasi. Beberapa library utama yang digunakan dalam penelitian ini antara lain:

1. NumPy

Library fundamental untuk komputasi numerik yang menyediakan fungsi untuk operasi array multidimensi dan matriks [9]. NumPy digunakan dalam penelitian ini untuk manipulasi data numerik, seperti perhitungan matriks kovarian dan operasi matematis lainnya yang menjadi dasar dalam implementasi PCA.

2. Pandas

Library untuk manipulasi dan analisis data struktural. Pandas memungkinkan peneliti untuk mengimpor, membersihkan, dan mengatur dataset dalam format tabel (dataframe) dengan efisien. Fasilitas ini mempermudah proses preprocessing data, seperti imputasi nilai yang hilang dan standarisasi.

### 3. Scikit-learn

Library pembelajaran mesin yang menyediakan alat untuk preprocessing data, implementasi algoritma pembelajaran mesin, serta evaluasi model. Dalam penelitian ini, Scikit-learn digunakan untuk menerapkan PCA, adaptive weighting, dan algoritma deteksi anomali seperti Isolation Forest dan DBSCAN [10].

### 4. Matplotlib dan Seaborn

Library visualisasi data yang digunakan untuk membuat grafik dan diagram yang membantu dalam eksplorasi data dan penyajian hasil analisis. Seaborn, sebagai ekstensi dari Matplotlib, mempermudah pembuatan visualisasi statistik yang informatif, seperti heatmap matriks kovarian dan distribusi nilai eigen.

### 5. TensorFlow

Framework pembelajaran mendalam yang digunakan untuk implementasi Autoencoders. TensorFlow dipilih karena fleksibilitasnya dalam membangun, melatih, dan mengoptimalkan jaringan saraf, yang sangat penting untuk mendeteksi anomali berbasis rekonstruksi data.

Proses eksperimen dilakukan pada platform Jupyter Notebook atau Google Collab. Jupyter Notebook menyediakan antarmuka interaktif untuk menulis, menjalankan, dan mendokumentasikan kode secara bersamaan. Google Collab dipilih untuk komputasi berbasis cloud yang memungkinkan penggunaan GPU secara gratis, sehingga mempercepat proses pelatihan model Autoencoders pada dataset yang besar. Kombinasi alat ini memberikan fleksibilitas dan efisiensi dalam setiap tahap penelitian, mulai dari preprocessing hingga evaluasi hasil.

### III.5 Kerangka Kerja Penelitian

Kerangka kerja penelitian ini dirancang untuk memberikan gambaran menyeluruh tentang alur penelitian dari awal hingga akhir. Setiap tahap dalam kerangka kerja disusun secara sistematis untuk memastikan bahwa penelitian berjalan dengan terarah dan menghasilkan temuan yang valid. Berikut adalah langkah-langkah utama dalam kerangka kerja penelitian ini:

#### 1. Pengumpulan Dataset

Tahap ini melibatkan pemilihan dan pengunduhan dataset multivariat dari repositori publik atau sumber terpercaya lainnya. Dataset dipilih berdasarkan kriteria yang telah ditentukan, seperti jumlah fitur yang tinggi, keberadaan ground truth untuk evaluasi, dan relevansi dataset terhadap tujuan penelitian.

#### 2. Preprocessing Data

Data mentah yang telah dikumpulkan melalui proses preprocessing untuk memastikan kualitas dan konsistensinya. Preprocessing meliputi standarisasi fitur, imputasi nilai yang hilang, dan penghapusan outlier jika diperlukan. Proses ini bertujuan untuk meningkatkan integritas data sebelum diterapkan pada model.

#### 3. Penerapan PCA dengan Adaptive Weighting

PCA diterapkan untuk mereduksi dimensi data dengan memilih komponen utama berdasarkan explained variance ratio tertentu. Teknik adaptive weighting diterapkan untuk memberikan bobot yang relevan pada fitur berdasarkan kontribusinya terhadap deteksi anomali. Hasil dari tahap ini adalah dataset yang telah direduksi dimensinya namun tetap mempertahankan informasi yang relevan.

#### 4. Penerapan Algoritma Deteksi Anomali

Dataset hasil transformasi digunakan sebagai input untuk algoritma deteksi anomali, seperti Isolation Forest, DBSCAN, dan Autoencoders. Setiap algoritma dikonfigurasi dan diuji untuk memastikan hasil deteksi yang optimal pada dataset.

## 5. Evaluasi Model

Tahap evaluasi mencakup pengukuran performa model menggunakan metrik precision, recall, dan F1-score, serta analisis efisiensi komputasi berdasarkan waktu eksekusi. Hasil evaluasi digunakan untuk menilai keunggulan pendekatan yang diusulkan dibandingkan metode tradisional.

Kerangka kerja penelitian ini digambarkan dalam diagram alur yang mengilustrasikan hubungan antar tahap, mulai dari pengumpulan data hingga evaluasi hasil. Diagram ini memberikan panduan yang jelas tentang bagaimana setiap langkah saling berhubungan dan mendukung tujuan penelitian secara keseluruhan.

### III.6 Batasan Penelitian

Penelitian ini memiliki beberapa batasan yang perlu diperhatikan dalam interpretasi hasil dan penerapan model pada skenario dunia nyata. Berikut adalah batasan utama penelitian ini:

#### 1. Struktur Dataset

Dataset yang digunakan dalam penelitian ini terbatas pada data multivariat dengan struktur linier. PCA, sebagai teknik utama dalam reduksi dimensi, bekerja paling baik pada data dengan hubungan linier antar fitur. Untuk dataset dengan hubungan non-linier, hasil PCA mungkin tidak sepenuhnya optimal tanpa penyesuaian lebih lanjut.

#### 2. Teknik Adaptive Weighting

Pembobotan adaptif dalam penelitian ini didasarkan pada relevansi fitur terhadap variabel target atau pola anomali. Pendekatan ini terbatas pada teknik berbasis korelasi dan informasi gain, sehingga tidak mencakup teknik pembobotan yang lebih kompleks, seperti berbasis pembelajaran mesin.

#### 3. Lingkungan Eksperimen

Penelitian dilakukan dalam lingkungan simulasi menggunakan dataset yang tersedia secara publik. Hasil penelitian mungkin berbeda jika diterapkan pada sistem dunia nyata yang

memiliki tantangan tambahan, seperti distribusi data yang lebih dinamis atau perubahan pola anomali.

#### 4. Skalabilitas Model

Meskipun efisiensi komputasi diukur dalam penelitian ini, skalabilitas model pada dataset dengan volume data yang sangat besar atau dengan fitur-fitur berdimensi tinggi tetap menjadi tantangan yang perlu diuji lebih lanjut.

#### 5. Keterbatasan Alat

Alat dan software yang digunakan, seperti Jupyter Notebook atau Google Colab, memiliki keterbatasan dalam hal daya komputasi, terutama saat menangani dataset yang sangat besar atau saat melatih model jaringan saraf yang kompleks seperti Autoencoders.

Batasan-batasan ini tidak hanya memberikan konteks untuk interpretasi hasil, tetapi juga membuka peluang untuk penelitian lebih lanjut yang dapat mengatasi keterbatasan tersebut dan memperluas aplikasi model dalam berbagai domain.

## Daftar Pustaka

- [1] Aggarwal, C. C. (2017). *Outlier Analysis* (2nd ed.). Springer.
- [2] Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics.
- [3] K Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [4] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- [5] Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: Identifying Density-Based Local Outliers. *Proceedings of the ACM SIGMOD International Conference on Management of Data*.
- [6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [7] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- [8] Hawkins, D. M. (1980). *Identification of Outliers*. Chapman and Hall.
- [9] Van der Maaten, L., & Hinton, G. (2008). Visualizing Data Using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.
- [10] Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.