



✧

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

Alya Izzah Zalfa R (21081010187)

✧

Pendahuluan

Latar Belakang

Data multivariat berdimensi tinggi sering kali mengandung redundansi informasi, yang memperumit analisis dan meningkatkan risiko overfitting. PCA dikenal efektif untuk reduksi dimensi, tetapi tidak mempertimbangkan relevansi fitur terhadap tujuan analisis.

Tujuan Penelitian

Mengintegrasikan PCA dengan adaptive weighting untuk menghasilkan reduksi dimensi yang efisien dan relevan, serta meningkatkan akurasi deteksi anomali.



Research Gap

Penelitian Sebelumnya

- Research on Image Recognition of Tomato Leaf Diseases Based on Improved AlexNet Model (Jing Qiu et al., 2024)
 - Fokus pada fusi berbobot fitur HOG dan LBP untuk meningkatkan akurasi pengenalan citra.
 - Relevansi: Pembobotan berbasis relevansi diterapkan, tetapi tidak mencakup integrasi PCA untuk reduksi dimensi.

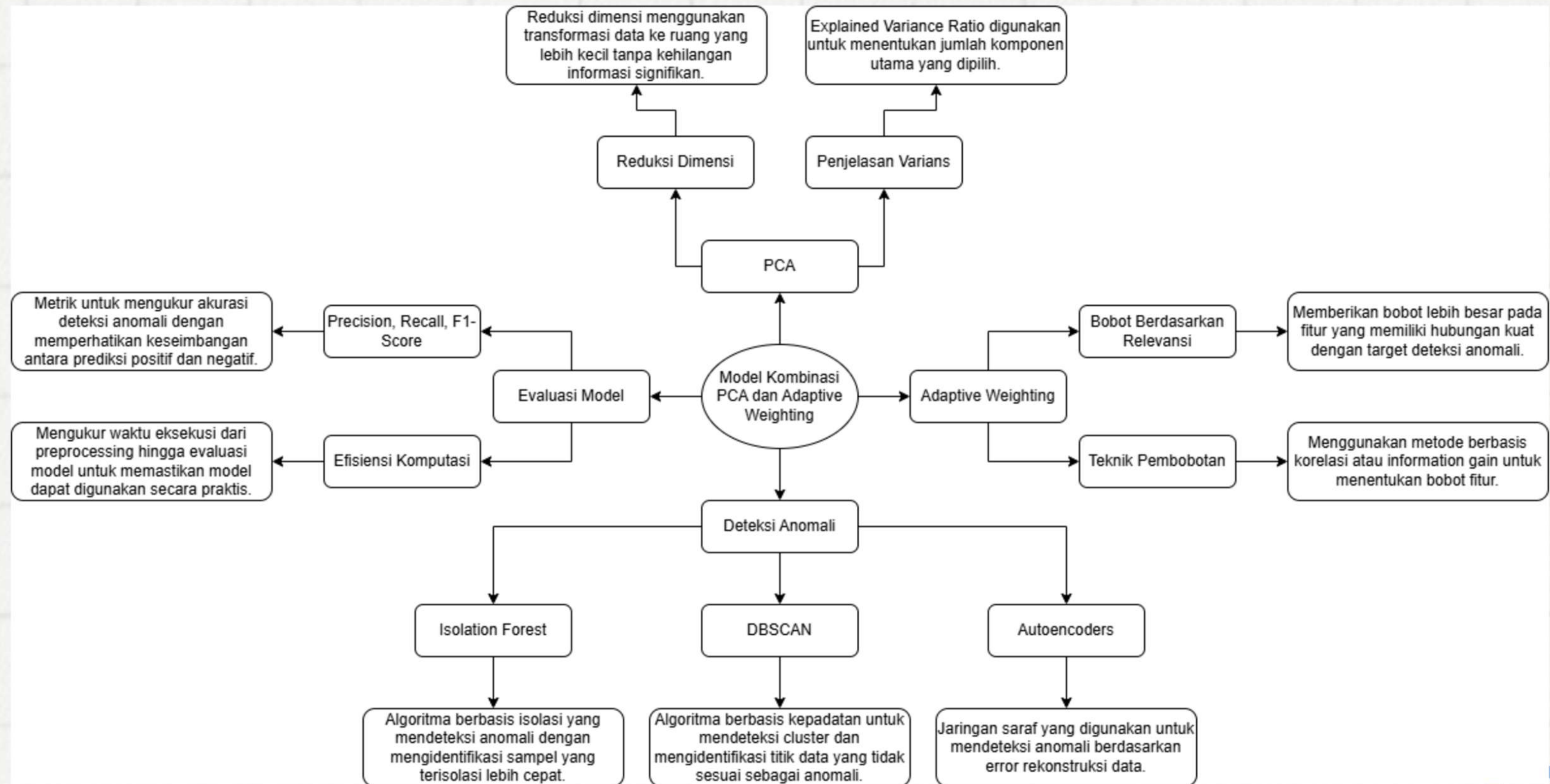
Research Gap

- PCA hanya mempertimbangkan varians data tanpa menilai relevansi fitur tertentu terhadap tujuan analisis.
- Penelitian sebelumnya memisahkan reduksi dimensi dan pembobotan fitur, yang menghasilkan representasi data kurang informatif.
- Dampak hasil reduksi dimensi terhadap algoritma deteksi anomali jarang dibahas secara mendalam.

Pendekatan baru yang mengintegrasikan PCA dan adaptive weighting untuk menghasilkan representasi data lebih relevan, meningkatkan akurasi deteksi anomali, dan efisiensi komputasi.



Mind map





Metodologi Penelitian

Metodologi yang digunakan adalah **kombinasi Principal Component Analysis (PCA)** dengan teknik Adaptive Weighting untuk reduksi dimensi dan deteksi anomali. PCA berperan sebagai alat utama untuk mengurangi dimensi data dengan memilih komponen utama yang menjelaskan sebagian besar varians data, sementara adaptive weighting digunakan untuk memberikan bobot lebih besar pada fitur yang relevan terhadap deteksi anomali.

Testing (Evaluation) Metrics

Precision

Precision digunakan untuk mengukur proporsi prediksi positif yang benar-benar merupakan anomali. Hal ini membantu dalam memahami tingkat akurasi deteksi terhadap kasus yang diidentifikasi sebagai anomali oleh model. Precision tinggi menunjukkan kemampuan model untuk meminimalkan kesalahan pada prediksi positif.

Recall

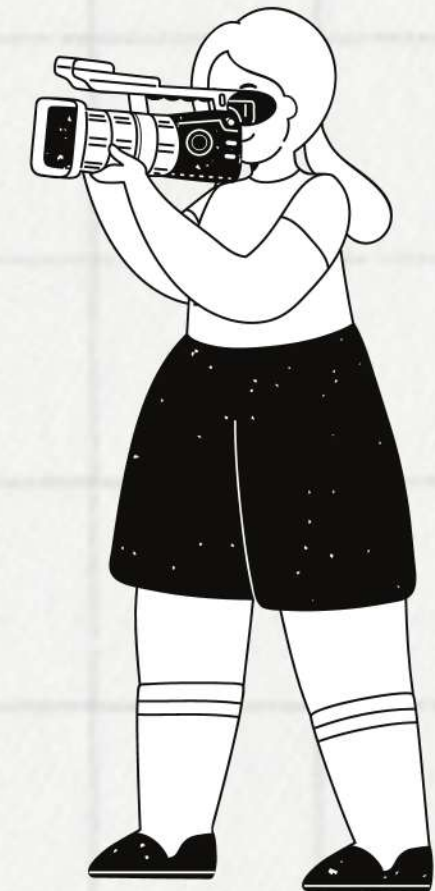
Recall mengukur kemampuan model untuk mendeteksi semua anomali dalam dataset. Recall yang tinggi menunjukkan bahwa model mampu menemukan sebagian besar anomali, termasuk yang sulit terdeteksi.

F1-Score

Merupakan rata-rata harmonis antara precision dan recall, memberikan keseimbangan antara keduanya, terutama pada dataset yang tidak seimbang. F1-Score sangat relevan ketika penting untuk mempertimbangkan trade-off antara precision dan recall.

Efisiensi Komputasi

Mengukur waktu eksekusi dari preprocessing, penerapan PCA dengan adaptive weighting, hingga deteksi anomali. Hal ini penting untuk memastikan bahwa model tidak hanya akurat, tetapi juga praktis. Evaluasi waktu dilakukan menggunakan data dalam berbagai ukuran dan tingkat kompleksitas, sehingga dapat diketahui skalabilitas model pada sistem besar.



Research Progress

1. Tahapan yang Telah Selesai:

- Pengumpulan dataset dari UCI Machine Learning Repository. Dataset yang digunakan mencakup data multivariat dengan jumlah fitur yang tinggi dan variasi distribusi.
- Preprocessing data melibatkan beberapa langkah: standarisasi data agar setiap fitur memiliki skala yang sebanding, imputasi nilai hilang menggunakan rata-rata atau median, dan penghapusan outlier untuk menjaga integritas analisis.

2. Tahapan Sedang Berjalan:

- Penerapan PCA untuk reduksi dimensi awal dilakukan dengan mempertahankan explained variance ratio sebesar 90–95%. Tahapan ini bertujuan untuk menyaring fitur yang relevan tanpa kehilangan informasi penting.
- Perhitungan bobot adaptif menggunakan metode korelasi Pearson dan mutual information untuk memberikan penekanan pada fitur yang paling relevan dengan target deteksi anomali.
- Integrasi adaptive weighting dengan PCA sedang diuji untuk menghasilkan representasi data yang optimal.

3. Tahapan Berikutnya:

- Pengujian algoritma deteksi anomali: Isolation Forest, DBSCAN, dan Autoencoders akan diuji menggunakan dataset hasil transformasi. Setiap algoritma dikonfigurasi untuk mengoptimalkan deteksi anomali sesuai karakteristik dataset.
- Evaluasi model dilakukan dengan menggunakan metrik precision, recall, F1-Score, dan efisiensi komputasi untuk memastikan kinerja model.



**Thank you
very much!**

