

PRAKTIKUM DATA MINING

Nama : Hawa Andini Hadi
NIM : 23051214238
Kelas/Angkatan : DTMG1/2023
Algoritma : DBSCAN dan PCA
Jenis Analisis : Anomaly Detection
Dataset : Estimation of Obesity Levels Based On Eating Habits and Physical Condition

Keterangan Dataset :

1. Gender: Jenis Kelamin
2. Age: Umur
3. Height: Tinggi
4. Weight: Berat
5. Family_history_with_overweight: Apakah ada anggota keluarga yang menderita atau menderita kelebihan berat badan?
6. FAVC: Apakah Anda sering mengonsumsi makanan berkalori tinggi?
7. FCVC: Apakah Anda biasanya mengonsumsi sayuran dalam makanan Anda?
8. NCP: Berapa banyak makanan utama yang Anda konsumsi setiap hari?
9. CAEC: Apakah Anda makan makanan apa pun di antara waktu makan?
10. SMOKE: Apakah kamu merokok?
11. CH2O: Berapa banyak air yang Anda minum setiap hari?
12. SCC: Apakah Anda memantau kalori yang Anda makan setiap hari?
13. FAF: Seberapa sering Anda melakukan aktivitas fisik?
14. TUE: Berapa banyak waktu yang Anda gunakan pada perangkat teknologi seperti telepon seluler, permainan video, televisi, komputer dan lainnya?
15. CALC: Seberapa sering Anda minum alkohol?
16. MTRANS: Transportasi apa yang biasanya Anda gunakan?
17. NObeyesdad: Tingkat obesitas

Pembahasan : Dataset ini berisi data tentang obesitas berdasarkan berbagai faktor seperti usia, berat badan, tinggi badan, kebiasaan makan, dan pola aktivitas. Variabel target adalah NObeyesdad, yang menunjukkan kategori obesitas seseorang.

Metode Preprocessing :

1. Pemilihan Kolom: Dari seluruh data, hanya dua kolom yang digunakan yaitu FCVC (seberapa sering makan sayur) dan CH2O (seberapa sering minum air).
2. Menghapus Data Kosong: Baris data yang kosong atau tidak lengkap dihapus supaya analisis lebih akurat.

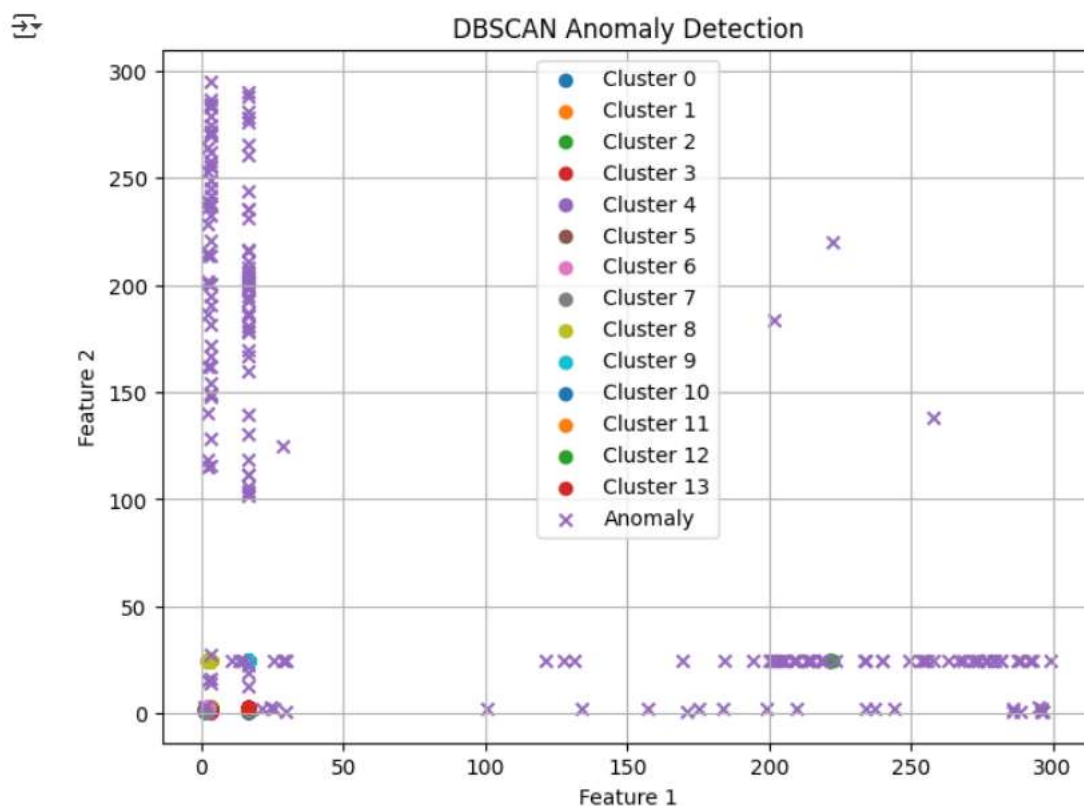
3. Menyesuaikan Skala Data: Nilai-nilai pada data disesuaikan agar memiliki skala yang sama, supaya analisis lebih seimbang dan tidak berat sebelah.

Metode Evaluasi Model :

1. DBSCAN: Metode ini digunakan untuk melihat apakah ada data yang "menyimpang" atau berbeda dari yang lain. Ia bekerja dengan mencari kelompok data yang padat, dan yang jauh dari kelompok itu dianggap aneh (anomali).
2. PCA: Metode ini mengubah data menjadi bentuk baru yang lebih sederhana, lalu dilihat seberapa jauh setiap titik dari pusat. Titik-titik yang sangat jauh dianggap sebagai data yang tidak biasa (anomali).
3. Visualisasi: Hasil dari kedua metode ditampilkan lewat grafik supaya lebih mudah dilihat dan dipahami.

Hasil Evaluasi :

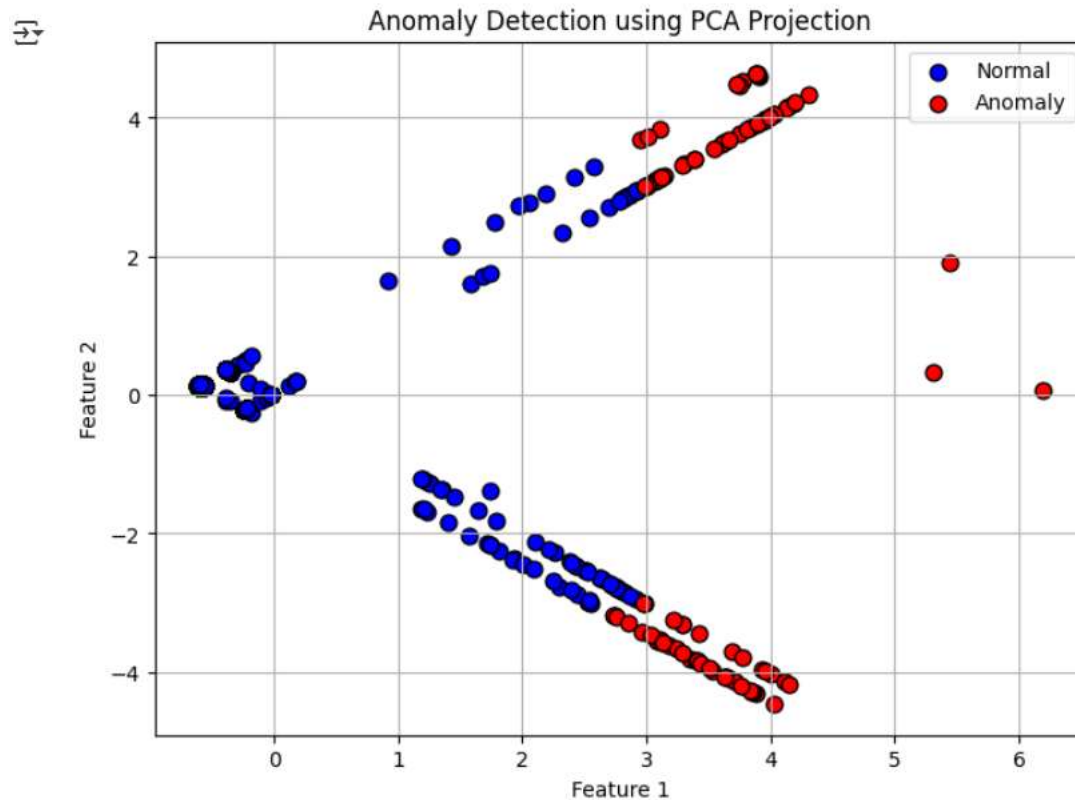
1. DBSCAN



Grafik ini menampilkan hasil pemisahan data menggunakan metode DBSCAN, di mana data dikelompokkan menjadi 14 klaster dan beberapa titik dikenali sebagai data tidak biasa atau anomali (ditandai dengan warna ungu dan simbol silang). Sebagian besar data berada di wilayah dengan nilai rendah, sedangkan anomali tersebar di area yang lebih jauh. Hal ini menunjukkan bahwa DBSCAN mampu menemukan pola umum dalam data serta mengenali

data yang berbeda dari pola tersebut. Meskipun begitu, jumlah kluster yang cukup banyak bisa jadi menandakan bahwa pengaturan parameter DBSCAN masih perlu disesuaikan.

2. PCA



Gambar ini menunjukkan hasil pendeteksian data yang tidak biasa menggunakan metode PCA, di mana data diubah menjadi dua komponen utama untuk memudahkan analisis. Titik berwarna biru merepresentasikan data yang dianggap normal, sedangkan titik merah menunjukkan data yang menyimpang atau anomali. Dari tampilan grafik, terlihat bahwa data anomali cenderung berada jauh dari kumpulan data utama, yang berarti karakteristiknya cukup berbeda. Secara keseluruhan, pendekatan ini mampu membedakan data normal dan anomali dengan cukup jelas berdasarkan pola penyebarannya.

Kesimpulan :

1. Baik DBSCAN maupun PCA dapat digunakan untuk mendeteksi data yang tidak biasa (Anomali).
2. DBSCAN lebih cocok digunakan jika data membentuk kelompok yang jelas, sementara PCA lebih efektif untuk melihat sejauh mana data berbeda dari nilai rata-rata.
3. Grafik yang dihasilkan dari kedua metode ini sangat membantu dalam menunjukkan hasilnya.

Lampiran:

1. Link data set dan google colab: https://colab.research.google.com/drive/1ZabAGG-unu_rX0hyzyyJBajrcWjiMlAs?usp=sharing