

Tugas Praktikum: Clustering (Advanced)

Deskripsi Tugas

Mahasiswa diminta mengimplementasikan K-Means, Gaussian Mixture Models (GMM), DBSCAN, dan Spectral Clustering pada Customer Marketing Dataset (People, Products, Promotion, Place). Tujuan praktikum adalah memahami perbedaan algoritma dan mengevaluasi hasil clustering dengan metrik yang sesuai untuk menyusun segmen pelanggan.

Tujuan Praktikum

1. Memahami konsep dan implementasi K-Means, GMM, DBSCAN, dan Spectral.
2. Membandingkan hasil clustering lintas algoritma.
3. Mengevaluasi performa menggunakan Silhouette Score (opsional: Calinski-Harabasz/Davies-Bouldin) dan menginterpretasi segmen.

Langkah-Langkah Praktikum

1. Dataset
Gunakan Customer Marketing Dataset https://github.com/kcv-if/Modul-PD/blob/master/Materi/6%20-%20Dataset_Tugas.csv
2. Eksplorasi & Pra-proses
 - o Tampilkan struktur & tipe data, cek *missing/outlier*.
 - o Konversi tanggal (Dt_Customer), buat fitur ringkas (mis. Age, Tenure, TotalSpend, Frequency = total purchases, PromoAcceptedCount).
 - o One-hot untuk kategori (Education, Marital_Status) dan standarisasi fitur numerik (z-score/RobustScaler).
3. Implementasi Clustering
 - o K-Means: cari k (2-10) dengan Elbow/Silhouette.
 - o GMM: bandingkan jumlah komponen & tipe kovarians via BIC/AIC.
 - o DBSCAN: tentukan eps (k-distance plot) & min_samples, catat noise.
 - o Spectral: pilih n_clusters, uji affinity (rbf/nearest_neighbors).
4. Visualisasi Hasil
 - o Scatter PCA-2D berwarna label klaster untuk tiap algoritma.
 - o Ringkas profil klaster (rata-rata z-score fitur kunci: Recency, Frequency, Monetary, channel, promo).
5. Evaluasi Clustering
 - o Hitung Silhouette untuk setiap algoritma (opsional: CH/DB).
 - o Jika ada label acuan (tidak wajib), hitung Rand Index/NMI.
 - o Bandingkan hasil dan jelaskan algoritma paling sesuai untuk dataset.
6. Laporan Praktikum
 - o Ringkas langkah, visualisasi, tabel metrik, dan profil segmen yang dapat ditindaklanjuti.
 - o Kesimpulan: algoritma terbaik & alasan (stabilitas, interpretabilitas, kualitas metrik).

Hasil yang Diharapkan

1. Implementasi keempat algoritma dengan visualisasi klaster (PCA-2D).
2. Tabel perbandingan metrik (min. Silhouette) dan ringkasan profil segmen.
3. Laporan singkat berisi hasil & analisis rekomendasi segmen.

Penilaian

1. Eksplorasi & Pra-proses (20%) — kebersihan data, encoding, scaling, fitur.
2. Implementasi Algoritma (40%) — penerapan & *tuning* K-Means, GMM, DBSCAN, Spectral.
3. Evaluasi & Analisis (30%) — metrik, perbandingan, dan interpretasi segmen.
4. Laporan (10%) — kerapuhan, kelengkapan, dan kejelasan analisis.