

Tugas Praktikum: Clustering (Advanced)

Deskripsi Tugas

Mahasiswa diminta untuk mengimplementasikan algoritma clustering lanjutan seperti DBSCAN, Gaussian Mixture Models (GMM), atau Spectral Clustering pada dataset. Praktikum ini bertujuan untuk memahami perbedaan algoritma clustering serta mengevaluasi hasil clustering menggunakan metrik yang sesuai.

Tujuan Praktikum

1. Memahami konsep dan implementasi algoritma clustering lanjutan.
2. Membandingkan hasil clustering dengan algoritma yang berbeda.
3. Mengevaluasi performa clustering menggunakan metrik seperti Silhouette Score dan Rand Index.

Langkah-Langkah Praktikum

1. **Unduh Dataset**

Gunakan dataset Mall Customer Segmentation Data dari Kaggle. Dataset ini berisi data pelanggan seperti usia, pendapatan tahunan, dan skor pengeluaran.

Link dataset: <https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python>

2. **Eksplorasi Data**

- Tampilkan struktur dataset dan atribut yang tersedia.
- Bersihkan data dari nilai kosong atau tidak valid.
- Standarisasi data numerik menggunakan z-score atau min-max scaling.

3. **Implementasi Clustering**

- Terapkan algoritma DBSCAN untuk menemukan kluster berdasarkan kepadatan.
- Terapkan Gaussian Mixture Models (GMM) untuk membagi data menjadi kluster berdasarkan distribusi.
- Terapkan Spectral Clustering untuk mempartisi data menjadi kluster.

4. **Visualisasi Hasil Clustering**

- Buat scatter plot untuk memvisualisasikan kluster hasil algoritma.
- Gunakan warna berbeda untuk setiap kluster.

5. **Evaluasi Clustering**

- Hitung metrik evaluasi seperti Silhouette Score untuk setiap algoritma.
- Jika tersedia label sebenarnya, hitung Rand Index atau Normalized Mutual Information (NMI).

- Bandingkan hasil dan interpretasikan algoritma mana yang lebih cocok untuk dataset tersebut.

6. ****Laporan Praktikum****

- Sertakan langkah-langkah, hasil visualisasi, dan evaluasi hasil clustering.
- Tulis kesimpulan tentang algoritma yang paling efektif untuk dataset ini dan alasan pemilihannya.

Hasil yang Diharapkan

1. Implementasi algoritma clustering lanjutan dengan hasil visualisasi kluster.
2. Perbandingan evaluasi algoritma menggunakan metrik seperti Silhouette Score dan Rand Index.
3. Laporan singkat yang mencakup hasil dan analisis.

Penilaian

1. ****Eksplorasi Data (20%)****: Pemahaman tentang struktur dan persiapan data.
2. ****Implementasi Algoritma (40%)****: Kesesuaian penerapan algoritma clustering.
3. ****Evaluasi Hasil Clustering (30%)****: Penggunaan metrik evaluasi yang tepat dan interpretasi hasil.
4. ****Laporan Praktikum (10%)****: Kerapihan, kelengkapan, dan analisis hasil.