

IF4074 Pembelajaran Mesin Lanjut

Spesifikasi Tugas Besar 2

Clustering Algorithm

Deskripsi Tugas

- Implementasikan lah algoritma *clustering* sebagai berikut
 1. DBSCAN
 2. *Agglomerative Hierarchical*
 3. K-Means
- Inisialisasi *centroid* awal pada algoritma K-Means dilakukan secara random
- Parameter yang dapat diubah dalam algoritma *clustering* yang akan diimplementasi adalah sebagai berikut:
 - Minpts pada algoritma DBSCAN
 - Epsilon pada algoritma DBSCAN
 - N-cluster pada algoritma K-Means dan *Agglomerative*
 - Jenis *linkage* (Single, Complete, Average, Average-Group Linkage) pada algoritma *Agglomerative*
- Perhitungan jarak menggunakan *euclidean*
- Fungsi clustering hanya menangani atribut bertipe numerik.
- Implementasi dilakukan dengan bahasa *python*
- Tugas dikerjakan secara berkelompok dengan maksimal 3 mahasiswa dalam 1 kelompok, dan harus berasal dari kelas yang sama.

Pengujian

- Setelah diimplementasikan, lakukanlah pengujian hasil implementasi terhadap *dataset iris* <<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>> dengan menghilangkan label.
- *Split* dataset iris menjadi *training* dan *test*. Lakukan proses *clustering* menggunakan data *training*. Lakukan pengujian dengan memprediksi data *test* terhadap *cluster* yang terbentuk dari data *training* oleh masing-masing algoritma yang diimplementasi.
- Evaluasi akurasi hasil *clustering* dari masing-masing algoritma yang diimplementasi. Akurasi dihitung dengan membandingkan label yang diprediksi oleh algoritma terhadap label sebenarnya.
- Lakukan juga proses *training* dan *test* dengan menggunakan algoritma yang terdapat pada *library* scikit-learn.

- Bandingkan akurasi algoritma yang diimplementasi dengan algoritma yang terdapat pada *library* scikit-learn.

Pengumpulan

Tugas dikumpulkan maksimal pada hari Senin, 18 November 2019 melalui situs kuliah online dalam sebuah file .zip yang berisi:

1. Kode program (dalam *file .py*)
2. Laporan (dalam *file .pdf*) yang berisi:
 - a. Penjelasan singkat terkait konsep masing-masing algoritma clustering
 - b. Penjelasan kode program (penjelasan singkat mengenai kelas, fungsi, dan variabel yang dibuat)
 - c. Hasil clustering *dataset iris* dengan menggunakan *classifier* yang dibuat
 - d. Analisis hasil eksperimen
 - e. Pembagian tugas dalam kelompok

Catatan:

Pertanyaan mengenai tugas ini dapat dituliskan pada *link* berikut <https://docs.google.com/spreadsheets/d/1ppOTDLmRy-ZsHkyIISKUVxbPI5RYK9qphHpneXvoWFw/edit?usp=sharing>