

IF4071 Pembelajaran Mesin

Tugas Kecil I: Eksplorasi scikit-learn untuk Clustering pada Jupyter Notebook

20 September 2018

Jupyter Notebook (<http://jupyter.org/>) memudahkan kita untuk membuat dan men-share dokumen yang merupakan gabungan dari live code, equation, visualisasi dan catatan. Jupyter dapat digunakan untuk visualisasi, pembersihan dan data transformasi, statistical model dan machine learning. Scikit-learn merupakan library machine learning pada bahasa python.

1. Lakukan eksplorasi scikit learn pada Jupiter Netbook dan bacalah dokumentasinya: <http://scikit-learn.org/stable/documentation.html>
2. Tulislah script dalam bahasa python pada satu notebook untuk melakukan task berikut ini:
 - a. Membaca dataset standar iris dan dataset play-tennis (dataset eksternal dalam format csv). Gunakanlah sklearn.datasets untuk membaca dataset standar. Untuk membaca dataset csv, gunakanlah Python Data Analysis Library <http://pandas.pydata.org/>
 - b. Melakukan eksplorasi scikit learn clustering Kmeans, Agglomerative clustering, DBSCAN, Gaussian mixtures terkait parameter yang bisa digunakan.
 - c. Untuk k-medoids, Graph-based clustering, dan grid clustering, eksplorasi library yang dapat digunakan.
 - d. Menyimpan (save) model/hipotesis hasil pembelajaran ke sebuah file eksternal
 - e. Membaca (read) model/hipotesis dari file eksternal
 - f. Membuat instance baru dengan memberi nilai untuk setiap atribut
 - g. Melakukan assignment cluster dengan memanfaatkan model/hipotesis clustering yang ada pada instance pada g.
4. Tugas dikerjakan per orang.
5. Tugas dikumpulkan berupa file .ipynb dan pdf dengan isi laporan sesuai ketentuan butir tiga.
6. Penamaan file yang dikumpulkan: Tucil1_[NIM].zip (misal: Tucil1_13515001.zip yang berisi Tucil1_13515001.ipynb dan Tucil1_13515001.pdf).
7. Pengumpulan yang terlambat tidak diperbolehkan, batas akhir adalah hari Kamis, 27 September 2018 jam 23.59 (waktu situs ini).