

Spesifikasi Tugas Kecil 1

Eksplorasi library Algoritme Pembelajaran pada Jupyter Notebook

Dikumpulkan: Senin, 27 Februari 2023 jam 08.59 WIB

Jupyter Notebook (<http://jupyter.org/>) memudahkan kita untuk membuat dan berbagi 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.

Lakukan eksplorasi Scikit-learn pada Jupyter Notebook dan bacalah dokumentasinya:

Jupyter notebook: <https://jupyter-notebook.readthedocs.io/en/stable/notebook.html>

Scikit-learn: <http://scikit-learn.org/stable/documentation.html>

Dataset eksplorasi: Breast-cancer (dataset internal sklearn)

Tuliskan script dalam bahasa python pada satu notebook untuk melakukan task berikut ini, untuk setiap dataset lakukan:

1. Membaca dataset (load dataset).

Gunakanlah sklearn.datasets untuk membaca dataset breast-cancer.

Membagi dataset menjadi 80% data training dan 20% data testing.

2. Melakukan pembelajaran dengan algoritma berikut:

- a. DecisionTreeClassifier (<http://scikit-learn.org/stable/modules/tree.html>),

- b. Id3Estimator (<https://github.com/svaante/decision-tree-id3>)

- c. K Means

(<https://scikit-learn.org/0.19/modules/generated/sklearn.cluster.KMeans.html>)

- d. LogisticRegression

(https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

- e. Neural_network

(https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html)

- f. SVM

(<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC>)

dengan memanggil method: `fit(datatrain, targettrain)` atau

`fit(datatrain)` untuk semua data training.

Anda dapat melakukan sedikit eksperimen untuk setting parameter method `fit()`.

Tuliskan parameter yang digunakan.

Catatan:

- Jika diperlukan encoding data kategorikal, gunakanlah library `LabelEncoder`.
- Khusus untuk `DecisionTreeClassifier`, tampilkan model pohon hasil `Decision Tree` dengan method `export_text`.

3. Simpan model hasil pembelajaran dengan pustaka `pickle`.
4. Melakukan proses prediction dengan load model yang sudah disimpan memanfaatkan pustaka `pickle`, kemudian memanggil method `predict(datatest)`. Lakukan evaluasi hasil prediksi dengan menggunakan metric: Accuracy, Precision, Recall dan F1; serta `confusion_matrix` hasil prediksi model.
5. Analisis hasil metrik evaluasi yang sudah diperoleh untuk semua algoritma pembelajaran, dalam bentuk perbandingan nilai dan penjelasan singkat mengenai hasil tersebut.
6. Lakukan k-fold cross validation (dengan k=10), dengan fungsi `cross_validate` pada algoritme `DecisionTreeClassifier` saja untuk dataset breast cancer. Tampilkan hasil akurasi dan F1 score dari 10-fold cross validation. Bandingkan dan analisis hasil perbandingan dengan akurasi dan F1 score `DecisionTreeClassifier` pada hasil di butir (4).

Tugas Kecil 1 ini dikerjakan berkelompok, terdiri atas 2 mahasiswa (boleh gabungan mahasiswa K1, K2, dan K3). Tugas dikumpulkan berupa hasil download notebook dalam dua format yaitu file `.ipynb` dan `pdf`. Tuliskan identitas semua anggota kelompok di dalam notebook tsb. Salah satu anggota kelompok mengunggah berkas tugas pada website kuliah ini. Pengumpulan di Edunex, tidak diperkenankan terlambat.

Penamaan file yang dikumpulkan: `Tucil_[NIM anggota1]_[NIM anggota2].zip`. (NIM Urut dari terkecil).

Batas akhir adalah hari Senin, 27 Februari 2023 jam 08.59.