

Spesifikasi Tugas Kecil 1

Eksplorasi library Decision Tree Learning pada Jupyter Notebook

Dikumpulkan: Minggu, 20 Februari 2022 jam 23.59

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:

1. Breast-cancer (dataset internal sklearn)
2. Play-tennis (dataset eksternal dalam format csv)

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.
Untuk membaca dataset csv, gunakanlah Python Data Analysis Library (pandas) <http://pandas.pydata.org/>
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(data_train, target_train)` atau `fit(data_train)` untuk semua data (full training), dan menampilkan modelnya.
Anda dapat melakukan sedikit eksperimen untuk setting parameter method `fit()`.

Tuliskan parameter yang digunakan.

Catatan:

- Jika diperlukan encoding data kategorikal, gunakanlah library `LabelEncoder`.
- Untuk menampilkan model pohonnya, gunakan method `export_text`.

Lakukan evaluasi model dengan menggunakan metric: Accuracy dan F1.

3. Melakukan proses prediction dengan memanggil method `predict(datatest)`.
Lakukan evaluasi hasil prediksi dengan menggunakan metric: Accuracy dan F1.
4. Analisis hasil Accuracy dan F1 yang sudah diperoleh untuk semua algoritma pembelajaran, dalam bentuk perbandingan nilai dan penjelasan singkat mengenai hasil tersebut.

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.

Setiap anggota kelompok me-*upload* file tugas pada website kuliah ini.

Penamaan file yang dikumpulkan: Tucil1_[NIM anggota1]_[NIM anggota2].zip.
(NIM Urut dari terkecil)

Pengumpulan di Edunex, tidak diperkenankan terlambat.

Batas akhir adalah hari Minggu, 20 Februari 2022 jam 23.59.