

Algorytmy analizy danych  
Kierunek Informatyka, WE, sem. VII  
Laboratorium nr 6

Grzegorz Cichosz

1. Cześć 1 poleceń

- a. Zapoznaj się ze skryptem języka Python zawartym w p6.txt
- b. Otwórz nowy projekt w środowisku PyCharm i utwórz nowy plik programu Python (nadaj mu nazwę np.  
p6 – czyli plik będzie się nazywał p6.py)
- c. Skopiuj skrypt do p6.py zawartość z pliku p6.txt
- d. Uruchom skrypt na wybranych danych (uwaga! musisz wskazać odpowiedni plik). Skorzystaj z danych z poprzednich zajęć.
- e. Wykonaj polecenia 1 i 2 zapisane w skrypcie
- f. Skrypt z wykonanymi poleceniami wyślij iliasem w odpowiednim do tego interfejsie.

Ponownie robię pythona w shellu

```
#zmiana nazwy  
mv p4.txt p4.py
```

```
#nadaj uprawnienia do uruchomienia  
chmod +x p4.py
```

```
#robimy virtualne środowisko  
python3 -m venv venv
```

```
#aktywujemy środowisko  
source venv/bin/activate
```

```
#instalujemy zależności  
pip install pandas seaborn matplotlib scipy kneed scikit-learn yellowbrick
```

#odpalamy

```
(venv) talandar@gentoo ~/workspace/aad/lab6 $ ./p6.py
Traceback (most recent call last):
  File "/home/talandar/workspace/aad/lab6/./p6.py", line 13, in <module>
    from yellowbrick.cluster import SilhouetteVisualizer
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/yellowbrick/__init__.py", line 31, in <module>
    from .style.rcmod import reset_defaults, reset_orig
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/yellowbrick/style/__init__.py", line 20, in <module>
    from .colors import *
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/yellowbrick/style/colors.py", line 33, in <module>
    from distutils.version import LooseVersion
ModuleNotFoundError: No module named 'distutils'
(venv) talandar@gentoo ~/workspace/aad/lab6 $ pip install distutils
ERROR: Could not find a version that satisfies the requirement distutils (from versions: none)
```

I mamy problem

Od Pythona 3.12 moduł distutils został usunięty

komentuję tą linię I próbuję jeszcze raz

from yellowbrick.cluster import SilhouetteVisualizer

znowu problem

program oczekuje iris? Ale nie ma w plikach na iliasie

próbuję użyć plików jakie są w zakładce lab6

df = pd.read\_csv('dane\_iris.csv', header=0)

```
(venv) talandar@gentoo ~/workspace/aad/lab6 $ ./p6.py
Tyle wierszy ma null: 0
Traceback (most recent call last):
  File "/home/talandar/workspace/aad/lab6/./p6.py", line 25, in <module>
    std_df = scaler.fit_transform(df)
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/sklearn/utils/_set_output.py", line 316, in wrapped
    data_to_wrap = f(self, X, *args, **kwargs)
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/sklearn/base.py", line 907, in fit_transform
    return self.fit(X, **fit_params).transform(X)
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/sklearn/preprocessing/_data.py", line 924, in fit
    return self.partial_fit(X, y, sample_weight)
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/sklearn/base.py", line 1336, in wrapper
    return fit_method(estimator, *args, **kwargs)
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/sklearn/preprocessing/_data.py", line 961, in partial_fit
    X = validate_data(
        self,
    ...<4 lines>...
        reset=first_call,
    )
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/sklearn/utils/validation.py", line 2902, in validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/sklearn/utils/validation.py", line 1022, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/sklearn/utils/_array_api.py", line 878, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/home/talandar/workspace/aad/lab4/venv/lib/python3.13/site-packages/pandas/core/generic.py", line 2171, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '14.08;1176.04;6967.05;1976.14;3090.08;4005.18;5063.12;6059.04'
```

dopytam na zajęciach co dalej.