

LABORATORIUM 3

Stosowanie klasyfikatorów uczenia maszynowego

Zestawy danych

<https://data.gov/> to portal otwartych danych rządu Stanów Zjednoczonych.

<https://www.ncdc.noaa.gov/cag/> - zapewnia portal klimatyczny NOAA

<https://www.esrl.noaa.gov/psd/data/timeseries/> - Portal badań Ziemi administracja laboratoriów (ESRL, Earth System Research Laboratory) NOAA dostarcza miesięczne i sezonowe z danymi klimatycznymi.

<https://www.quandl.com/search> - Quandl udostępnia setki bezpłatnych szeregów czasowych z danymi finansowymi, a także zestawy danych z płatnym dostępem.

<https://datamarket.com/data/list/?q=provider:tsdl> - biblioteka danych TSDL (biblioteka danych szeregów czasowych) zawiera linki do setek tymczasowych zbiorów danych rządu w wielu obszarach przemysłowych.

<http://archive.ics.uci.edu/ml/datasets.html> - repozytorium uczenia maszynowego Uniwersytetu Kalifornijskiego w Irvine (UCI) zawiera dziesiątki zestawów danych szeregów czasowych z różnych obszarów.

ZADANIE

1. Należy napisać program (za pomocą biblioteki **scikit-learn**), który realizuje następujące metody(algorytmy):

- regresji logistycznej
- maszyny wektorów nośnych
- jądra SVM
- Uczenia drzew decyzyjnych
- Algorytm k-najbliższych sąsiadów

Proszę o wybranie jednego zestawu danych (dowolnego, które Państwa interesują) i sprawdzenie na nim algorytmów klasyfikacji.

Na ocenę 3 proszę o zaprogramowanie dowolnego (jednego!) algorytmu,

na 4 dowolne dwa algorytmy,

na 5 dowolnych trzech algorytmów

Zrób podsumowanie: który z algorytmów uzyskał lepsze wyniki w klasyfikacji, dlaczego