

# Scikit-learning

niedziela, 1 listopada 2020

12:42

-train\_test\_split - dzieli zbiór na dwa podzbiory (treningowy i testowy) z możliwością zdefiniowania zarodka liczby losowej. Może przekazywać jej wiele zbiorów o tej samej liczbie kolumn - będą one rozdzielone między takie same indeksy.

-StratifiedShuffleSplit - służy do dzielenia zbioru z użyciem metody losowania warstwowego

-Imputer klasa której użyć można do przekształcania danych w zbiorze. Do konstruktora dorzucamy strategię której imputer będzie używał np. strategy='mean'

-imputer.fit(DataFrame) = dopasowanie obiektu imputer do danych. Działa jedynie na dane numeryczne, więc wcześniej należy wyzbyć się danych kategoryalnych i tekstowych

-imputer.transform(DataFrame) - zwraca transformowane dane z DataFrameu

-imputer.statistics\_ - są tu przechowywane wartości median lub inne (w zależności od strategii) zbioru, z którym wywołaliśmy metodę fit

-LabelEncoder() - klasa do kodowania atrybutów numerycznych

-encoder.fit\_transform(Data\_set) - dokonuje transformacji

-OneHotEncoder() - służy do transformacji gorącojedynkowej

-one\_hot\_encoder.fit\_transform(encoded\_data.reshape(-1,1)) - w rezultacie mamy **macierz rzadką** (jeśli chcemy **gęstą** to używamy **toarray()**)

-LabelBinarizer().fit\_transform(data\_frame) - tworzy macierz gęstą (jeśli chcemy rzadką to do konstruktora dajemy **sparse\_output=True**)

-Pipeline() - pozwala na stworzenie własnego potoku przekształcającego

-FeatureUnion() - pozwala na łączenie potoków

-corss\_val\_score() - Pozwala na przeprowadzenie walidacji krzyżowej

-joblib - pozwala na zapis i odczyt wytrenowanych modeli

-GridSearchCV() - służy do przeprowadzania regularyzacji modelu z użyciem metody przeszukiwania siatki