

Laborator 6

Modele de clasificare

Folositi urmatoarele seturi de date:

1. [Semeion Handwritten Digit Data Set \(http://archive.ics.uci.edu/ml/datasets/Semeion+Handwritten+Digit\)](http://archive.ics.uci.edu/ml/datasets/Semeion+Handwritten+Digit)
2. [Wireless Indoor Localization Data Set \(http://archive.ics.uci.edu/ml/datasets/Wireless+Indoor+Localization\)](http://archive.ics.uci.edu/ml/datasets/Wireless+Indoor+Localization)
3. [Spambase Data Set \(http://archive.ics.uci.edu/ml/datasets/Spambase\)](http://archive.ics.uci.edu/ml/datasets/Spambase)
4. [Smartphone Dataset for Human Activity Recognition \(HAR\) in Ambient Assisted Living \(AAL\) Data Set \(http://archive.ics.uci.edu/ml/datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+](http://archive.ics.uci.edu/ml/datasets/Smartphone+Dataset+for+Human+Activity+Recognition+%28HAR%29+)

plus inca doua seturi de date de clasificare alese de voi, din repository-urile specificate in Cursul 4.

1. Daca e necesar, aplicati o metoda de missing value imputation sau eliminati inregistrarile/coloanele care au valori lipsa; documentati metoda folosita.
2. Pentru fiecare set de date aplicati 5 modele de clasificare din scikit learn. Pentru fiecare raportati: acuratete, scor F1 - a se vedea [sklearn.metrics \(http://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics\)](http://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics) folosind 5 fold cross validation. Raportati mediile rezultatelor atat pentru fold-urile de antrenare, cat si pentru cele de testare. Revedeti formele ulterioare ale acestui document pentru precizari despre: continutul rezultatelor raportate, modalitate de notare.
3. Documentati in jupyter notebook fiecare din modelele folosite, in limba romana. Puteti face o sectiune o sectiune separata cu documentarea algoritmilor + trimitere la algoritm.
4. Pentru fiecare model: efectuati o cautare a hiperparametrilor optimi folosind grid search si random search.

Pentru fiecare set de date raportati rezultatele obtinute de fiecare model.



Predare: laboratorul va fi salvata in repository-ul de github si predat in saptamana 15-19 aprilie.