

# Laborator 6

## Modele de clasificare

Folositi 5 seturi de date pentru probleme de clasificare, plecand de la repository-urile specificate in Cursul 6. Toate seturile de date trebuie sa aiba valori precizate (adica sa fie fara valori lipsa) si sa aiba macar o trasatura de intrare variabila categoriala nominala.

1. Transformati trasaturile categoriale nominale folosind one hot encoding, [https://pandas.pydata.org/docs/reference/api/pandas.get\\_dummies.html](https://pandas.pydata.org/docs/reference/api/pandas.get_dummies.html).
2. (numar de modele *numar de seturi de date* \ 1 punct = 30 de puncte) Pentru fiecare set de date aplicati 6 modele de clasificare din scikit learn. Pentru fiecare raportati: acuratete, precision, recall, scorul F1 - a se vedea [sklearn.metrics](#), [Precision and recall](#) - folosind 10 fold cross validation. Raportati mediile rezultatelor atat pentru fold-urile de antrenare, cat si pentru cele de testare. Rularile se vor face cu valori fixate ale hiperparametrilor.
3. (numar de modele *numar de seturi de date* 1 punct = 30 de puncte) Raportati performanta fiecarui model, folosind 10 fold cross validation. Pentru fiecare din cele 10 rulari, cautati hiperparametrii optimi folosind 4-fold cross validation. Performanta modelului va fi raportata ca medie a celor 10 rulari. *Observatie:* la fiecare din cele 10 rulari, hiperparametrii optimi pot diferi, din cauza datelor utilizate pentru antrenare/validare.
4. (numar modele \* 4 puncte = 20 puncte) Documentati in jupyter notebook fiecare din modelele folosite, in limba romana. Daca acelasi algoritm e folosit pentru mai multe seturi de date, puteti face o sectiune separata cu documentarea algoritmilor + trimitere la algoritm.

Se acorda 20 de puncte din oficiu.

Exemple de modele de clasificare:

1. [Multi-layer Perceptron classifier](#)
2. [KNN](#)
3. [SVM](#)
4. [Gaussian processes](#)
5. [RBF](#)
6. [Decision tree](#)
7. [Random forest](#)
8. [Gaussian Naive Bayes](#)

*Predare:*

1. Predarea se face cel tarziu in 25 noiembrie 2022 ora 23, in lucrarea de pe elearning (Tema 5).
2. Obligativu: type annotations pentru variabile, parametri, tip de retur; docstrings.

3. Fișierele de date folosite vor fi descarcate local de studenți și puse într-un director "data".  
Se va realiza o arhivă zip care conține minim: fișierul/fișierele ipynb și directorul de date.  
Suplimentar, pot fi folosite imagini incluse în ipynb; acestea vor fi puse în directorul "images" ce se va include în arhivă zip predata.