

ZADANIE PPREDIKCIA

Cieľom zadanie je navrhnúť model pre predikciu/klasifikáciu dát. Je možné využiť modely pre klasifikáciu/regresiu dostupné v rámci modulu scikit-learn.

Maximálny počet bodov: **15 b** (prípadne podľa pokynov na cvičeniach).

Podrobnejšie info na cvičeniach a na prednáške.

DÁTA:

K dispozícii su dáta X_{public} o rozmere N vzoriek, kde každá vzorka obsahuje F príznakov, teda $N \times F$. K týmto dátam je zverejnený odpovedajúci vektor značiek y_{public} o rozmere N prvkov. Dáta je možné načítať pomocou `numpy.load()`.

Overenie vášho modelu bude realizované na dátach X_{eval} rozmeru $N_e \times F$, ktoré sú generované rovnakým kľúčom ako X_{public} .

Odobzdvávanie vektora predikovaných značiek, ktorý vygeneruje váš klasifikátor na dátach X_{eval} je 27.11.2021 o 16:00.

HODNOTENIE:

Študent odovzdáva vektor predpovedí y_{eval} o rozmere N_e prvkov. Tento je získaný aplikovaním navrhnutého modelu na dáta X_{eval} .

Úspešnosť predikcie bude vyhodnotená mierkou `roc_auc_score`

Na základe dosiahnutej úspešnosti pri predikcii budú študenti zoradení do tabuľky začínajú študentom, ktorý dosiahne najvyššiu hodnotu `accuracy_score` na X_{eval} dátach. S klesajúcou úspešnosťou bude klesať počet bodov pridelených študentovi.

Omeškanie odovzdania zadania je penalizované odpočítaním 4b, za každý načatý deň omeškania.

TERMÍN:

27.11.2021 16:00

ČO JE POTREBNÉ ODOVZDAŤ:

1. Vektor predikovaných značiek o rozmere N_e prvkov, uložený pomocou príkazu `numpy.save()` v `numpy` formáte. Meno súboru je potrebné zadať **`y_predikcia.npy`** . **Nekomprimovať!!!**
2. Zdrojový kód k riešeniu (v prípade viacerých súborov zdrojového kódu, je potrebné skomprimovať do jedného balíka)
3. Krátku technickú správu **v pdf formáte** o rozsahu cca 1, maximálne 3 strany, pozostávajúcu z
 - teoretická časť - stručný teoretický opis použitého klasifikátora
 - návrh riešenia - metodológia riešenia, môže obsahovať zdôvodnenie, prečo bola ktorá metóda, prípadne funkcia použitá
 - diskusia a výsledky - môže obsahovať výsledky, ktoré sa vám podarilo dosiahnuť na tréningových dátach, porovnanie rôznych prístupov ktoré ste skúšali, závery, zhodnotenie a podobne.