

Projekt #2: Klasifikácia

Zadanie

Vašou úlohou je rozoznávať údaje vytvorené ľuďmi a údaje generované strojom. Ide teda o binárnu klasifikáciu.

Úplné zadanie ako aj údaje potrebné na vypracovanie sú dostupné na adrese:

<http://web.tuke.sk/fei-cit/wikt2017/challenge.html>

Dáta

Dátová sada obsahuje údaje vytvorené na základe veľmi známeho datasetu MNIST redukciou dimenzionality. Máte dostupnú redukovanú sadu údajov ako aj dekóder do pôvodnej dimenzionality. Tento dekóder slúži primárne ako prostriedok vizualizácie redukovaných údajov. Pravidlá súťaže nebránia použitiu tohto dekóderu na to aby ste transformovali údaje do pôvodného rozmeru. Pre potreby splnenia požiadaviek na tento projekt to ale nie je nutné a vzhľadom na výrazné predĺženie a skomplikovanie tréningu modelov to nie je ani odporúčané. Ak by ale niekto veľmi chcel, tak nech sa páči.

Súťaž

Okrem odovzdania projektu v rámci predmetu máte možnosť zapojiť sa do súťaže. Budem veľmi rád ak to čo najviac z vás aj spraví. Pravidlá súťaže sú na stránke výzvy.

Hodnotenie

Za projekt je možné získať maximálne **40b**. Celkové hodnotenie pozostáva z nasledovných čiastkových položiek:

- Prieskumná analýza - 6b
- Feature engineering - 6b
- Použitie viacerých štandardných klasifikátorov - 6b
- Učenie súborom metód - 6b
- Výber optimálneho modelu - 6b
- Zdokumentovanie - 5b
- Celkový dojem - 5b

Odovzdanie

Termín na odovzdanie zadania je do stredy 13.12.2017 23:59. Do miesta odovzdania v AISe odovzdajte .zip súbor so všetkými vašimi Jupyter notebookmi a .csv súbor s predikciou vášho najlepšieho modelu, vo formáte takom ako predpisuje Data science challenge, z ktorého používame údaje. Tento súbor bude použitý na vytvorenie interného poradia. Pozícia v tomto poradí nebude mať negatívny vplyv na hodnotenie. Jupyter notebooky by mali okrem kódu a výsledkov obsahovať aj komentáre k výsledkom a ich interpretáciu. Dajte si pozor, aby ste odovzdali notebooky spolu s vloženými výsledkami

a vykreslenými diagramami. Odporúčam jednotlivé hodnotené časti zadania odovzdávať v samostatných notebookoch.