

# **I-SUNS: Zadanie č.3**

Súborové učenie

**Vypracoval:** Jakub Šíp

**AIS ID:** 91419

Na implementáciu tohto zadania som si zvolil programovací jazyk Python 3.8 a knižnicu sklearn a tensorflow. Pre predspracovanie dát som použil knižnicu pandas, ktorá uľahčuje prácu s dátovými štruktúrami. Na

### Úloha 1:

- Dáta som len skontroloval na prítomnosť NaN hodnoty (nevyplnenej bunky). Žiadna ďalšia úprava nebola potrebná.

### Úloha 2:

- Implementoval som Boosting a Bagging algoritmus.
- Boosting algoritmus po vyladení dosahuje presnosť 95%. Použil som parametre `n_estimators=550` a `algorithm=SAMME`

```
-----BOOSTING-----  
[[930  48   9]  
 [ 83 882   1]  
 [  3   0 979]]  
0.9509369676320273
```

- Ako druhý som implementoval Bagging algoritmus, ktorý má presnosť len 71% a jeho tréning trvá hrozne dlho. Zmenou parametru `n_estimators` by sa to asi dalo zlepšiť, ale potom to trvá nekonečne dlho.

```
-----BAGGING-----  
[[862   1 124]  
 [205 617 144]  
 [335   0 647]]  
0.7243611584327087
```

- Okrem slabých som ešte implementoval RandomForest, ktorý pri `max_depth=50` dosahuje presnosť až 97,8%. Z použitých algoritmov je najpresnejší. Graf ako tento strom vyzerá je priložený, lebo do tohto dokumentu sa nezmestí.

```
-----RANDOM FOREST-----  
[[977   7   3]  
 [ 43 922   1]  
 [  2   0 980]]  
0.9809199318568995
```

- Boosting a RandomForest pracujú s približne rovnakou časovou náročnosťou a porovnateľnou presnosťou, ale Random Forest je ešte o pár % presnejší. Bagging algoritmus je pomalý a v približne rovnakom časovom intervale má presnosť len 71%, takže na tento príklad nie je moc efektívny.

### Úloha 3:

- Na regresiu som implementoval algoritmy Decision Tree Regression a Linear Regression.
- Decision Tree dosahuje v mojej implementácii presnejšie výsledky. S presnosťou 42% je omnoho presnejší ako lineárny regresor.

```
-----DECISION TREE REGRESSION-----
0.4225004846893657
```

- Linear Regressor aj po vyskúšaní všetkých možných kombinácií parametrov dosahuje presnosť len 17% a teda na použitie aspoň v tomto prípade nie je moc vhodný

```
-----LINEAR REGRESSION-----
0.17479218644712116
```

### Nepovinné úlohy:

- Stacking algoritmus po vyladení parametrov dosahuje v klasifikovaní veľmi vysokú presnosť 98% a jeho beh trvá zlomok času algoritmu Bagging. Presnosťou je porovnateľný s random forest algoritmom.

```
-----STACKING-----
[[970   9   8]
 [ 37 928   1]
 [  2   0 980]]
0.9805792163543441
```