

Seminarska naloga

## **Uporaba rezervoarskega vzorčenja v strojnem učenju**

Mentor: doc. dr. Matjaž Kukar ([matjaz.kukar@fri.uni-lj.si](mailto:matjaz.kukar@fri.uni-lj.si))

Velikost skupine: 2 študenta

Rezervoarsko vzorčenje je statistična tehnika, stara že več kot 30 let, ki je svoj prerod doživela v sodobni analizi podatkov in strojnem učenju s pojavom sproti zajemajočih podatkovnih zbirk – podatkovnih tokov (data streams). Ideja rezervoarskega vzorčenja je ustvariti vnaprej omejen končen reprezentativen vzorec neskončnega toka podatkov. Cilj seminarske naloge je ustvariti pregled uporabe tehnike rezervoarskega vzorčenja v strojnem učenju in morebitnih njihovih izpeljank, ki se tičejo specifične uporabe in lastnosti podatkovnih tokov (npr. concept drift). Končni izdelek naj bo poročilo v obliki znanstvenega članka (5-10 strani), ki vsebuje vse zgoraj navedene elemente in kratke napotke za uporabo tehnike.

### Literatura:

- Vitter, Jeffrey Scott. Random sampling with a reservoir. ACM Trans. Math. Softw., 11(1):37–57, 1985. (originalni članek)
- Peilin Zhao, Steven C.H. Hoi, Rong Jin, Tianbao Yang, Online AUC Maximization, Proc. ICML 2011 (primer letošnjega članka)
- Robert Kessl, Parallel algorithms for mining of frequent itemsets, PhD Thesis, 2011, [http://star.fit.cvut.cz/~kessler/phd\\_thesis.pdf](http://star.fit.cvut.cz/~kessler/phd_thesis.pdf) (uporaba za paralelizacijo algoritma iskanja pogostih skupaj nastopajočih predmetov)
- Drugi spletni viri