

Projekt z predmetu Strojové učenie

2017/2018

V rámci projektu pôjde o riešenie praktického problému pomocou algoritmov strojového učenia pomocou balíka [WEKA](#). Cieľom projektu je získať skúsenosti s aplikáciou týchto algoritmov.

Riešený problém: **konštrukcia klasifikátora** (vo forme rozhodovacieho stromu, booleovskej formuly, rozhodovacieho zoznamu, neurónovej siete a pod.) na základe tréningovej množiny. Riešený problém môže byť stanovený aj inak po dohode s vyučujúcim.

Vyberte si nejaké dáta (tréningovú a testovaciu množinu) a aspoň dva vhodné algoritmy implementované v balíku WEKA, použite vybrané algoritmy na konštrukciu klasifikátora na základe tréningovej množiny, spracujte výsledky a skúsenosti v tvare textu v rozsahu ≥ 5 (A4) strán, a odovzdajte v elektronickej podobe (PDF, Word). Popis riešenia problému má pozostávať z nasledujúcich častí:

- Popis dát: čo sú to za dáta, čo reprezentujú, z akého zdroja pochádzajú, aké majú vlastnosti (≤ 3 body)
- Popis použitých algoritmov: ako reprezentujú svoje hypotézy, ako fungujú (≤ 4 body)
- Popis evaluačných kritérií: ako a čím ste merali kvalitu získaných výsledkov - na osobitnej testovacej množine, pomocou cross-validation, atď. (≤ 3 body)
- Popis výsledkov: nájdené hypotézy a ich presnosť, porovnanie úspešnosti použitých algoritmov, prezentácia výsledkov pomocou tabuliek/grafov (≤ 5 bodov)
- Interpretácia výsledkov vlastnými slovami, porovnanie výsledkov s očakávaniami: prečo bol ten-ktorý algoritmus lepší/horší, diskusia (napr. porovnať zrozumiteľnosť nájdených riešení, výpočtovú rýchlosť jednotlivých algoritmov, návrhy na možné vylepšenia) (≤ 5 bodov)

Vyššie uvedené body budú udelené na základe kvality jednotlivých častí referátu. Za projekt možno získať 20 bodov.

Vhodné dáta (trénovacie a testovacie množiny) je možné nájsť napr. v [UCI Machine Learning Repository](#), ale je možné použiť aj iné zdroje.

Projekty budú prezentované na cvičeniach na konci semestra.

Predpokladáme, že v skupine každý študent bude pracovať s inými dátami.

Projekt je potrebné odovzdať **do 5. mája 2018 (21m do 23. 4. 2018)**.