

**Pokyny k vypracovaniu projektov
pre 2-INF-150: Strojové učenie, Leto 2009
Vyučujúci: Tomáš Vinař, Pavel Petrovič**

Základnou úlohou v rámci projektu je:

- vybrať si nejaký problém strojového učenia (môže byť takmer čokoľvek, napr. aj niečo čo potrebujete na diplomovú prácu),
- naimplementovať jednu alebo viacero metód alebo variantov riešenia tohto problému,
- otestovať implementáciu na skutočných alebo simulovaných dátach,
- vyhodnotiť úspešnosť zameraním výkonnosti na testovacích dátach a porovnaním buď s inou voľne dostupnou implementáciou, triviálnym algoritmom nevyužívajúcim pokročilé metódy (napr. ľahko implementovateľná heuristika), alebo viacero variant medzi sebou (napr. rôzne kernely v prípade SVM, alebo rôzne rozvoje báz pri regresii)

Projekt treba odovzdať najneskôr 168 hodín (týždeň) pred konaním skúšky; v projekte sa **hodnotí iba to, čo je napísané a odovzdané**. Odovzdávajú sa dve časti: **report o projekte v papierovej forme** na sekretariáte Katedry infomatiky a **zdrojové texty a dáta elektronicky** ako jeden .zip súbor e-mailom na adresu vinar@fmph.uniba.sk.

Požiadavky na report. Report má byť v rozsahu 5-15 strán a je hlavným podkladom pre hodnotenie projektu. Report musí obsahovať najmenej:

- Nadpis a meno autora projektu.
- Abstrakt (cca 1-2 odstavce) sumarizujúci cieľ, použité metódy a výsledky projektu.
- Úvod do problematiky: popis úlohy, ktorú riešite; odkiaľ ste získali dáta a ich stručnú charakteristiku (napr. veľkosť trénovacej / testovacej množiny, koľko pozitívnych / negatívnych príkladov, priemernú veľkosť inštancie, popis atribútov, priemery a štandardné odchýlky zaujímavých atribútov a pod.)
- Popis a zdôvodnenie použitých metód (aké metódy, prečo?)
- Stručné technické detaily implementácie (v akom jazyku, aké knižnice ste použili, s akými problémami ste sa stretli a ako ste ich riešili)
- Experimentálna evaluácia: akým spôsobom vyhodnocujete úspešnosť implementácie? Porovnanie s niečim iným (viď vyššie): tabuľky, grafy, ich interpretácia a závery.
- Záver: myslíte si že projekt bol úspešný alebo neúspešný? prečo? keby ste projekt robili znovu, čo by ste urobili ináč? aké iné problémy by sa ešte dali riešiť podobnými metódami? (Poznámka: aj dobre spracovaný neúspech s dobrou analýzou príčin neúspechu môže získať plný počet bodov.)

Inšpirovať sa môžete aj na nasledujúcej stránke (aj keď nejde o presne rovnaké zadanie): <http://www.stanford.edu/class/cs229/projects2006.html>

Požiadavky na elektronickú časť.

- Komentované zdrojové texty vlastných programov a skriptov,
- Dáta s ktorými ste pracovali (ak ide o verejne prístupné dáta z internetu, alebo sú dáta veľké a zip súbor by presiahol 5MB, tak môžete napísať namiesto toho www linku do readme.txt súboru),
- Súbor readme.txt popisujúci jednotlivé súbory, ako ich skompilovať a spustiť

Hodnotenie (orientačné):

- Výber problému: 5%
- Výber a správne použitie metód na riešenie: 30%
- Implementácia: 20%
- Dizajn experimentu: 25%
- Report: 20%

(Kvalita reportu vplýva aj na ostatné časti – ak napríklad z reportu nemožno pochopiť, o čo sa vlastne snažíte, tak to značne zníži známku aj v ostatných častiach.)

Opisovanie. Tak ako vo zvyšku kurzu, opisovanie (od spolužiakov, alebo z internetu) bude prísne postihnuté – vid'. formálne upozornenie na stránke predmetu. Keďže ide o veľkú časť známky, v prípade zistenia opisovania projektu nemožno rátať s úspešným ukončením predmetu.