

**Pokyny k vypracovaniu projektov**  
**2-INF-150: Strojové učenie, Zima 2012**  
**Vyučujúci: Tomáš Vinar**

Základnou úlohou v rámci projektu je:

- vybrať si nejaký problém, na ktorý sa dajú aplikovať metódy strojového učenia (môže byť takmer čokoľvek, obzvlášť vhodné je niečo, čo potrebujete urobiť do diplomovej práce),
- naimplementovať jednu alebo viacero variant riešenia problému,
- otestovať implementáciu na skutočných alebo simulovaných dátach,
- vyhodnotiť úspešnosť zmeraním výkonnosti na testovacích dátach a porovnaním buď s inou voľne dostupnou implementáciou, triviálnym riešením nevyužívajúcim pokročilé metódy (napr. ľahko implementovateľnou heuristikou), alebo porovnať viacero variant medzi sebou (napr. rôzne kernely pre SVM, rôzne rozvoje báz a pod.).

**Návrh projektu** je max. jedna strana textu, ktorý obsahuje popis problému, ktorý sa chystáte riešiť, dáta na ktorých sa chystáte svoj projekt trénovať a testovať, ako aj popis spôsobu vyhodnocovania experimentov. Nie je potrebné do návrhu písať, aké metódy sa chystáte použiť na samotné riešenie problému. Návrh projektu je potrebné **odovzdať do stanoveného termínu počas semestra**.

**Samotný projekt** treba odovzdať **do termínu určenom na web stránke**. Projekt je súčasťou hodnotenia za semester a hodnotí sa iba čo je napísané a odovzdané. Projekt treba odovzdať v dvoch častiach:

- **písomný report** v papierovej forme (pod dvere M163)
- **zdrojové texty a dáta** ako jeden zip súbor e-mailom na adresu [vinar@fmph.uniba.sk](mailto:vinar@fmph.uniba.sk)

**Písomný report** má byť v rozsahu 5-15 strán (nehodnotí sa dĺžka, ale obsah) 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, odkiaľ ste získali dáta a ich stručná charakteristika (napr. veľkosť trénovacej a testovacej množiny, koľko pozitívnych a negatívnych príkladov, priemernú veľkosť každej inštancie, popis atribútov, priemery a štandardné odchýlky najzaujímavejších atribútov a pod.).
- Popis a zdôvodnenie použitých metód.
- Stručný popis technických detailov implementácie (v akom jazyku, aké knižnice ste použili a prečo, s akými technickými problémami ste sa stretli pri riešení).
- Experimentálne vyhodnotenie: spôsob vyhodnocovania úspešnosti, porovnanie s niečím iným, tabuľky, grafy, ich interpretácia a závery.
- Záver: Myslíte si, že bol projekt úspešný alebo neúspešný? Prečo? Keby ste projekt robili znovu, čo by ste urobili ináč? Aké problémy by sa dali riešiť podobným spôsobom?

**Aj dobre spracovaný neúspešný projekt môže získať plný počet bodov**, ak obsahuje dobrú analýzu príčin neúspechu. (Úspešné riešenia v praxi obvykle vyžadujú niekoľko neúspešných pokusov nasledovaných podrobnou analýzou príčin neúspechu.)

## **Elektronická časť** by mala obsahovať

- 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 výsledný súbor by presiahol 5 MB, môžete napísať linku do súboru readme.txt).
- Súbor readme.txt popisujúce jednotlivé súbory, ako ich skompilovať a spustiť.

## **Orientačné hodnotenie:**

- Vhodnosť výberu problému: 5%
- Výber a správne použitie metód: 30%
- Implementácia: 20%
- Dizajn experimentu: 25%
- Kvalita reportu: 20%

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

**Opisovanie.** Tak ako vo zvyšku tohto kurzu, opisovanie (či už od spolužiakov alebo z internetu) bude prísne postihované – vid' formálne upozornenie na stránkach predmetu. Keďže ide o veľkú časť hodnotenia, v prípade zistenia opisovania nemožno rátať s úspešným ukončením predmetu. **Za opisovanie pokladáme aj ak odovzdáte prácu, ktorú ste už v podobnej forme odovzdali na iný predmet.**