

# Inštrukcie k projektu z predmetu Machine Learning

Cieľom projektu je vytvorenie ML modelu (alebo viacerých) na vybrané dáta a problém (klasifikácia, regresia). Dáta si zvolí študent sám - odporúča sa použiť knižnicu dát na [www.kaggle.com](https://www.kaggle.com), prípadne github alebo iné. Študent vypracuje projekt v programovacom jazyku Python – kompletný kód (napr. v Jupyteri) spustený a uložený ako súbor .html spolu s dokumentáciou (podrobný popis jednotlivých krokov projektu, napr. v .pdf, .pptx, markdown) pošle vyučujúcemu cez MS Forms – link bude prístupný v skupine Machine Learning v MS Teams. Vypracovanie vo dvojiciach.

## Projekt by mal obsahovať nasledovné časti:

1. Exploratívna analýza, spracovanie a príprava dát na modelovanie – zoznámenie sa s dátami, pochopenie premenných v datasete, pochopenie vzťahov medzi premennými, závislá premenná a nezávislé premenné (v prípade učenia s učiteľom), opisné štatistiky, grafická analýza, identifikácia chýbajúcich hodnôt a odlahlých pozorovaní, prípadné transformácie dát (škálovanie, kódovanie premenných, logaritmická transformácia) a iné.
2. Výber ML algoritmu a zdôvodnenie výberu, stručný popis vybraného algoritmu.
3. Rozdelenie datasetu na trénovaciu a testovaciu časť.
4. Aplikácia ML algoritmu a vyhodnotenie jeho výkonnosti. Prípadne aj porovnanie viacerých algoritmov a zdôvodnenie výberu najlepšieho z nich. Optimalizácia hyperparametrov.

**Hodnotenie (max 40 bodov):** počet bodov získa študent podľa náročnosti projektu a schopnosti odpovedať a reagovať na otázky vyučujúceho k prezentovanému projektu. V rámci projektu bude pozitívne hodnotené nielen preukázanie dosiahnutých znalostí počas semestra, ale tiež preukázanie nových znalostí v oblasti ML, ale hlavne jazyka Python, vrátane nových kódov alebo použitia nových modulov (knižníc) a podobne.

**Termín odovzdania:** najneskôr týždeň pred prezentáciou

**Prezentácia:** na cca 10 minút + 5 minút diskusia/otázky. Stručný opis jednotlivých krokov.