<https://www.geeksforgeeks.org/how-to-do-train-test-split-using-sklearn-in-python/>

<https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

**Inštrukcie k projektu z predmetu Machine Learning**

Témou projektu je použitie niektorého z ML algoritmu na vybrané dáta. Dáta si volí študent sám - odporúča sa použiť knižnicu dát na [www.kaggle.com](http://www.kaggle.com/), prípadne github alebo iné. Študent vypracuje projekt v programovacom jazyku Python – kompletný kód (napr. v Jupyteri) odovzdá spolu s projektom(dokumentáciou – podobný popis jednotlivých krokov projektu).

**Projekt by mal obsahovať nasledovné časti:**

1. Exploratívna analýza – 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 odľahlých pozorovaní, prípadné transformácie dát (škálovanie, kódovanie premenných, logaritmická transformácia).

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

**Termín odovzdania:** najneskôr týždeň pred prezentáciou

**Prezentácia:** na cca 10 minút. Stručný opis jednotlivých krokov