

BI-ML1.21 přednáška 1

Karel Klouda

FIT ČVUT

5. 10. 2023

Autoři: Karel Klouda, Daniel Vašata.
Problémy, návrhy apod. hlase v [GitLabu](#).
Verze souboru: 5. října 2023 10:00.

Vyučující předmětu

- **Karel Klouda** (přednáší, cvičí, právě hovoří),

- **Daniel Vašata** (přednáší, cvičí a garantuje)

- **Ondřej Tichý** (cvičí)

Podmínky získání zápočtu

- **Zápočet bude postaven na vypracovávání domácích úkolů.**
- Úkoly budou během semestru **dva**, každý za **max. 25 bodů**.
- Domácí úkoly budete vypracovávat v jazyce Python ve formátu Jupyter notebook (.ipynb).
- Zadání a podrobné instrukce k vypracování a odevzdání najdete na stránkách předmětu:

 courses.fit.cvut.cz/BI-ML1/ 

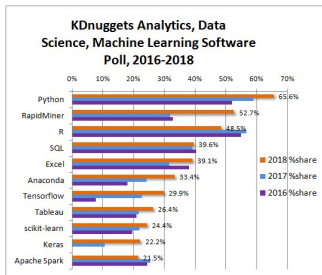
- **K získání zápočtu je třeba získat alespoň 25 bodů z 50.**

Podmínky složení zkoušky

- Zkouška bude **ústní**.
- Každý student dostane **dvě otázky z předem zveřejněného seznamu**.
- Z každé otázky můžete získat **až 25 bodů**.
- Celkem tedy můžete ze zkoušky získat **až 50 bodů**. Není žádný minimální nutný počet bodů, který je třeba ze zkoušky získat.
- Pokud student/ka u zkoušky prokáže zásadní neznalost, může zkoušející použít **právo veta a zkoušku ukončit jako neúspěšnou**.
- V případě úspěchu u zkoušky se výsledná známka odvodí ze součtu bodů ze semestru a ze zkoušky.

O čem to všechno vlastně bude?

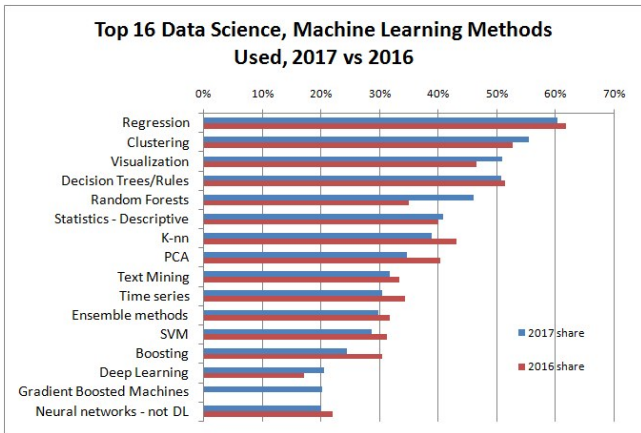
- Ve vší obecnosti: budeme se učit, jak z dat získat informace.
- Anglická *buzz words*, která popisují náplň tohoto předmětu, jsou **data mining** a **machine learning**; mezi těmito oblastmi je tenká a nežřetelná hranice.
- V tomto předmětu budeme používat jazyk **Python** a zejména balíčky, které se pro práci s daty používají.



Článek s novějšími daty.

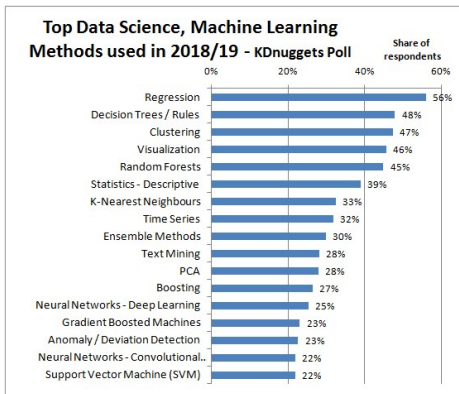
Metody a algoritmy (1/2)

- Cílem tohoto kurzu je naučit Vás základy dané problematiky, a tedy se zaměříme především na základní metody a algoritmy a na řešení problémů spojených s jejich použitím.
- Ovšem i úplně základní metody stále patří k nejpoužívanějším v praxi!



Metody a algoritmy (2/2)

- Cílem tohoto kurzu je naučit Vás základy dané problematiky, a tedy se zaměříme především na základní metody a algoritmy a na řešení problémů spojených s jejich použitím.
- Ovšem i úplně základní metody stále patří k nejpoužívanějším v praxi!



[zdroj: www.kdnuggets.com]

Z čeho se učit

- Studijní materiály tohoto předmětu by Vám měly stačit pro jeho absolvování: jsou to prezentace k přednáškám (vč. jejich handout formátu), Jupyter notebooky ke cvičením a svým způsobem i úkoly.
- Video z (letošních i dřívějších) přednášek a cvičení.
- Probírané metody patří ke klasické látce a lze k nim najít nepřeborné množství videí a textů.
- Pěkným zdrojem je [dokumentace](#) ke knihovně `scikit-learn`, kterou budeme pro aplikování probíraných modelů především používat.
- Užitečným zdrojem (nejen) zajímavých datasetů je server www.kaggle.com. Tam najdete uživateli vytvořené příklady použití různých metod, tipy a triky, soutěže (i o docela slušné peníze) atp.

Co už byste měli umět (1/2)

- Předpokládáme, že už umíte programovat, ale nepředpokládáme znalost Pythonu.
- Naučit se základy Pythonu, hlavně specializovaných knihoven, je součást tohoto kurzu.
- O některých metodách už jste slyšeli v *BI-ZNS: Znalostní systémy* (stará akreditace), ale fakticky o látce nemusíte vědět nic ;). Vše si zopakujeme a hlavně půjdeme více do hloubky.
- Předpokládáme znalost lineární algebry (*BI-LIN* resp. *BI-LA1* a *BI-LA2*) a matematické analýzy (*BI-ZMA* resp. *BI-MA1* a souběžné *BI-MA2*), neb tabulky jsou matice a strojově učit znamená optimalizovat!!!
- Také předpokládáme, že průběžně studujete předmět *BI-PST: Pravděpodobnost a statistika*.
- Budeme Vám muset říci alespoň trochu o hledání extrémů funkcí více proměnných (viz *BI-MA2*) a pár dalších věcí.

Co už byste měli umět (2/2)



Machine Learning in a nutshell

[zdroj: xkcd.com/1838/]

Navazující a související předměty (1/3)

- BI-ML2 Strojové učení 2:** Navazuje přímo na tento předmět, neuronové sítě, posilované učení a další techniky.
(učí Daniel Vašata, letní semestr).
- ?I-SZ Seminář znalostního inženýrství:** V rámci tohoto předmětu budete zpracovávat (Vámi) vybrané téma pod vedením jednoho z vyučujících či hostů.
(organizuje Magda Friedjungová, Pavel Kordík a Rodrigo Alves, letní i zimní semestr).
- BI-SVZ Strojové vidění a zpracování obrazu:** Práce s kamerovými systémy a zpracování obrazu z takových zařízení. Odehrává se v [Improlabu](#).
(přednáší doc. Marcel Jiřina a Improlab, zimní (a asi i letní) semestr).
- BI-VIZ Vizualizace dat:** Jak vizualizovat data, nástroje i postupy.
(přednáší Magda Friedjungová, zimní semestr).
- BI-PRS Praktická statistika:** Statistika nad reálnými daty, nástroj R.
(Petr Novák a Kamil Dedecius, letní semestr).

Navazující a související předměty (2/3)

NI-PDD Předzpracování dat: Zaměřeno na přípravu dat pro různé metody. Tato fáze je svým způsobem nejdůležitější a má největší vliv na kvalitu výsledku.
(přednáší doc. Marcel Jiřina, letní semestr, povinný pro ZI).

NI-UMI Umělá inteligence: Navazuje a rozšiřuje předmět *BI-ZUM*.
(přednáší doc. Pavel Surynek, letní semestr, povinný pro ZI).

NI-ADM Algoritmy data miningu Nejblíže příbuzný *BI-VZD*, na který přímo navazuje. Zaměření je podobné, témata ovšem pokročilejší.
(přednáší doc. Pavel Kordík, Karel Klouda a Daniel Vašata, zimní semestr, povinný pro ZI)

NI-MVI Metody výpočetní inteligence Je zaměřený především na moderní použití neuronových sítí.
(přednáší doc. Pavel Kordík a spol, letní semestr, povinný pro ZI)

Navazující a související předměty (3/3)

NI-SCR Statistická analýza časových řad: Předmět je zaměřen na práci s daty, které mají formát časové řady.

NI-BML Bayesovské metody ve strojovém učení: Praktické využití základních metod bayesovského modelování v dynamicky se rozvíjející oblasti machine learningu.
(přednáší Kamil Dedecius a Ondřej Tichý, letní semestr).

NI-GNN Grafové neuronové sítě: Umělá inteligence pro práci s grafovými daty.
(učí Miroslav Čepek, letní semestr).

NI-EDW Podnikové datové sklady: Prakticky zaměřený předmět o návrhu, architektuře a budování datových skladů.
(organizuje M. Friedjungová, letní semestr).

NIE-PML Personalised ML: Cílem kurzu je naučit studenty vytvářet modely na základě vlastností a chování jednotlivých entit (uživatelů).
(učí Rodrigo Alves, zimní semestr)

... ..