Výskumne orientovaný seminár 2

Mária Bieliková, Róbert Móro, Jakub Ševcech

25.9.2019

O nás...

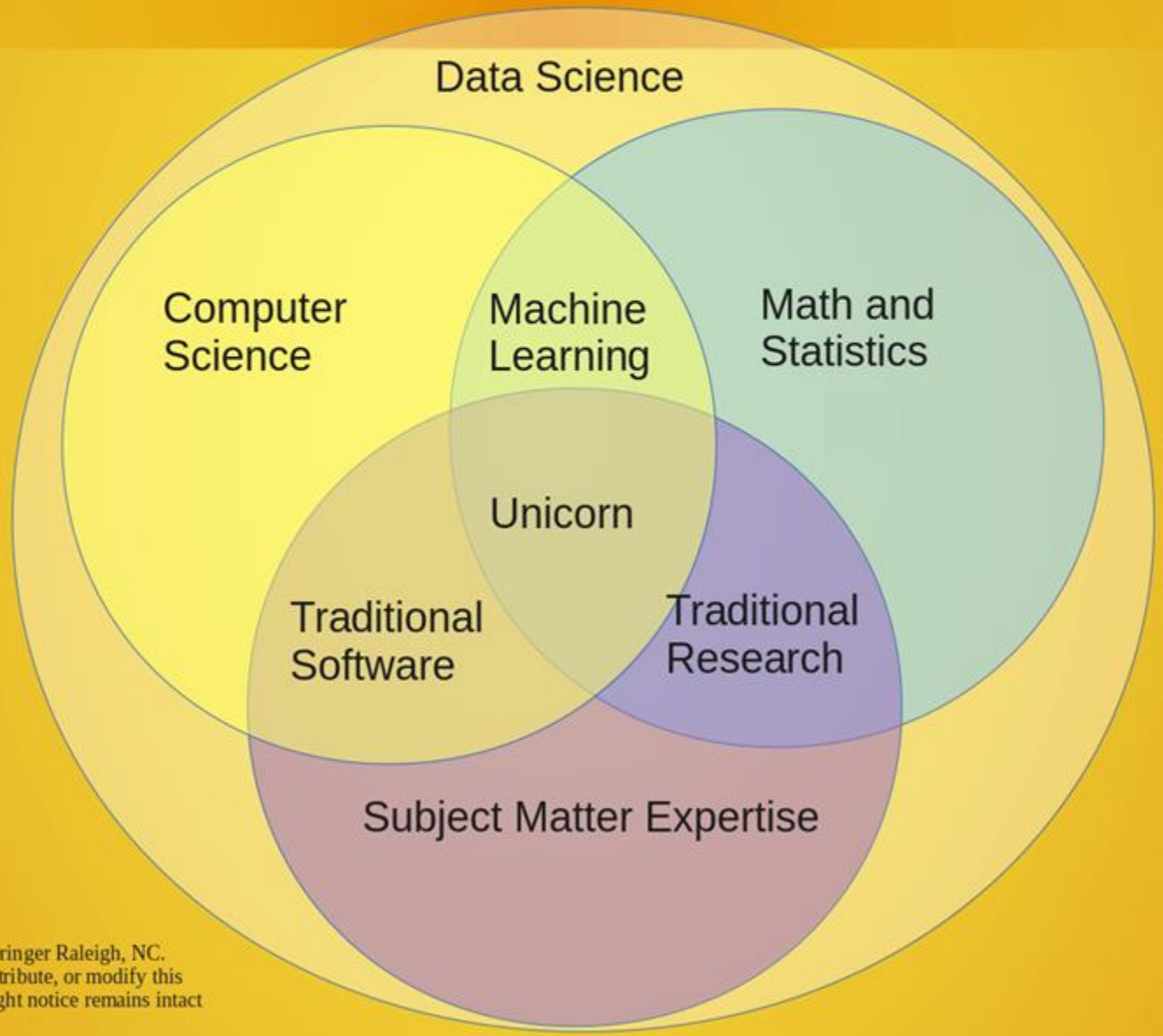
O vás...

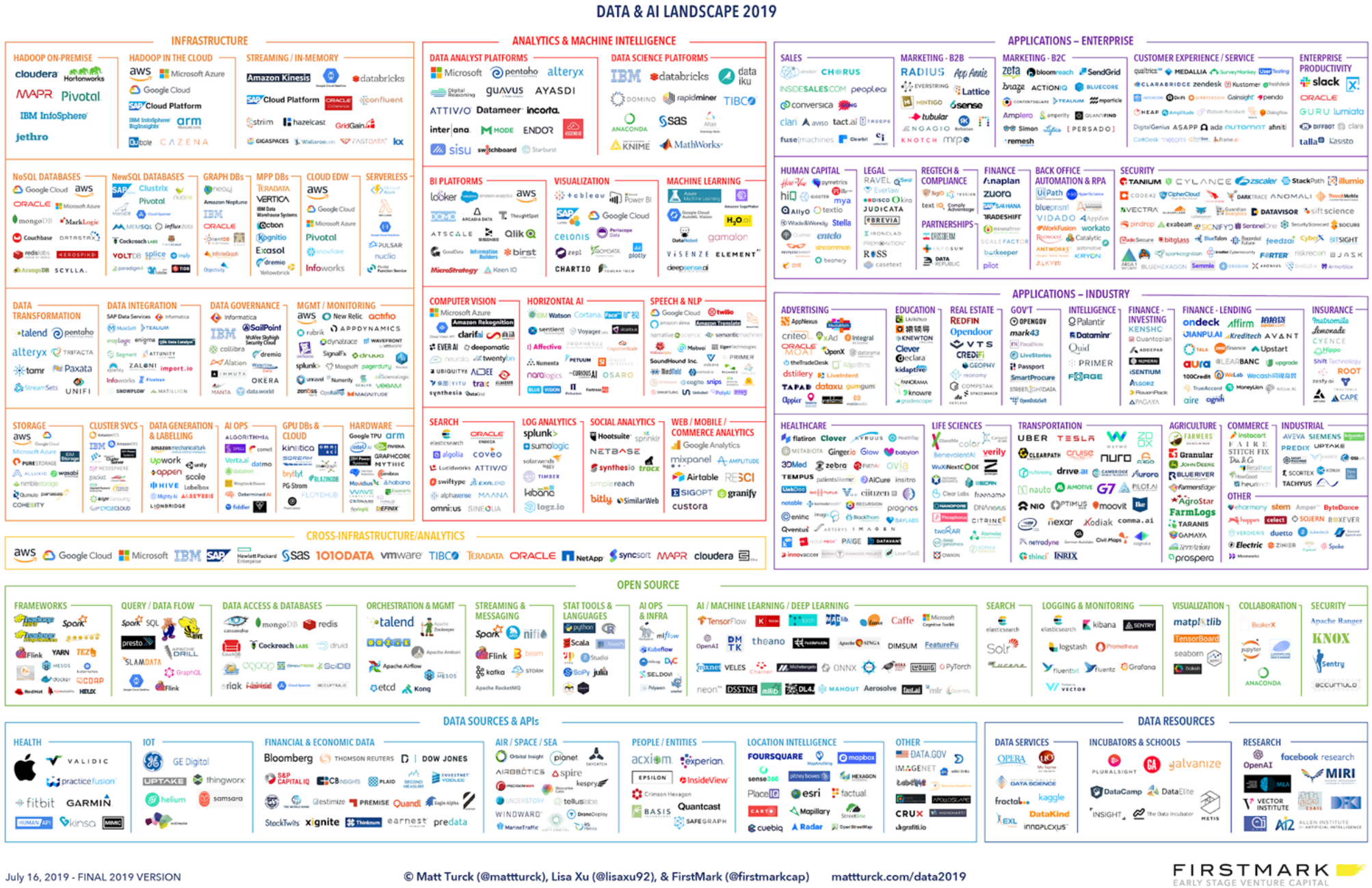
Na seminároch sa budeme venovať dátovej vede

Neexistuje presná definícia, čo je to dátová veda. Dátový vedec je...

- Sexy označenie pre štatistika
  - http://www.statisticsviews.com/details/feature/5133141/Nate-Silver-What-I-need-from-statisticians.html
- Programátor v oblasti strojového učenia / dolovania v dátach, ktorý má znalosť o doméne
  - http://shakthydoss.com/technical/what-is-the-difference-between-artificial-intelligence-machine-learning-statistics-and-data-mining/#comment-261
- Pokus o rebranding; niekto, kto ovláda štatistiku lepšie ako ktorýkoľvek softvérový inžinier a je lepší v softvérovom inžinierstve ako ktorýkoľvek štatistik
  - https://www.quora.com/What-is-data-science/answer/Rahul-Agarwal-10

Data Science Venn Diagram v2.0





https://mattturck.com/data2019/

Štatistika vs. strojové učenie vs. dolovanie v dátach

- http://shakthydoss.com/technical/what-is-the-difference-between-artificial-intelligence-machine-learning-statistics-and-data-mining/
- Ak opisujete dáta (deskriptívna štatistika) alebo odvádzate závery o populácii na základe jej vzorky (inferenčná štatistika), tak ste zrejme *štatistik*
- Ak využívate inferenčnú štatistiku na tvorbu algoritmov, ktoré sú schopné samé sa učiť, tak zrejme robíte *strojové učenie*
- Ak využívate strojové učenie na riešenie konkrétneho problému a deskriptívnu štatistiku na opísanie dát a výsledkov, tak zrejme *dolujete v dátach*

Čo robí dátový vedec?

1. Definuje (formuluje) otázky a hypotézy
2. Definuje ideálnu dátovú sadu
   a. Zisťuje, k akým dátam má prístup
   b. Získava dáta
   c. Čistí ich
   d. Skúma ich vlastnosti (exploratívna analýza)
3. Identifikuje a realizuje vhodný typ analýzy
   a. Štatistické modelovanie/predikcia
4. Interpretuje a komunikuje výsledky (dátový produkt)
5. Automatizuje kroky 2-4 pre predikcie na nových dátach

Úvodný kvíz na zamyslenie

1. „Keď chceš mať viac lajkov, musíš v statuse používať emotikony,“ tvrdí kamarát. Ako by ste jeho tvrdenie ako dátový vedec overili? Ako by ste pri tom argumentovali? Na čo si pritom treba dať pozor?
2. Vlastníte byt, ktorý chcete predať. Potrebujete teda určiť jeho cenu.
   a. Ako by ste ako dátový vedec postupovali?
   b. Aké dáta by ste na to potrebovali? Kde by ste ich získali?
3. Banka zbiera údaje o klientoch (rádovo tisíce klientov) a chcela by zistiť, ako na základe nich rozhodnúť, komu poskytnúť úver, a komu nie.
   a. Aké dáta by ste od bankára vyžadovali? V akej forme?
   b. Ako by ste ako dátový vedec postupovali pri návrhu spôsobu rozhodovania?
   c. Ako by ste bankárovi vysvetlili/prezentovali vami navrhnutý model?
   d. Ako by ste presviedčali banku, že je vami navrhnutý spôsob spoľahlivý?

Organizácia seminárov a podmienky absolvovania

VOS2 vs. IAU

Organizácia a témy seminárov

- Nástroje na analýzu a spracovanie dát
  - Python a R
- **Blok I: Základy analýzy dát a strojového učenia**
  - Exploratívna analýza a predspracovanie dát
  - Transformácie a tvorba odvodených čŕt
  - Tvorba a optimalizácia parametrov modelov strojového učenia
  - Vyhodnocovanie a výber modelov
  - Komunikovanie výsledkov
  - *Prípadová štúdia 1*
- **Blok II: Metódy racionality**
  - *Esej* o identifikovaných metódach racionality v zadanom štúdijnom texte (vypracovaná v 2-3 členných tímoch)
  - http://yudkowsky.net/rational, http://www.hpmor.com
- **Blok III: Pokročilé témy**
  - *Prípadová štúdia 2* na zadanú tému vypracovaná v 2-3 členných tímoch

Podmienky absolvovania

Klasifikovaný zápočet

Základy R	**10b**	(zadanie 2. týždeň, odovzdanie do seminára v 3. týždni)
Základy Pythonu	**10b**	(zadanie 2. týždeň, odovzdanie do seminára v 3. týždni)
Prípadová štúdia 1	**30b**	(zadanie v 3. týždni, odovzdanie v 7. týždni)
Analýza metód racionality	**10b**	(zadanie v 7. týždni, odovzdanie v 10. týždni)
Prípadová štúdia 2	**40b**	(zadanie v 7. týždni, záväzné nahlásenie témy v 9. týždni, odovzdanie v 12. týždni)
Spolu	**100b**	