

A vertical splash of water with many bubbles and droplets, set against a dark blue background. The water is captured in motion, creating a sense of freshness and energy.

Data Science projekty v telekomunikační společnosti

Jan Romportl

Chief Data Scientist, O2 Czech Republic

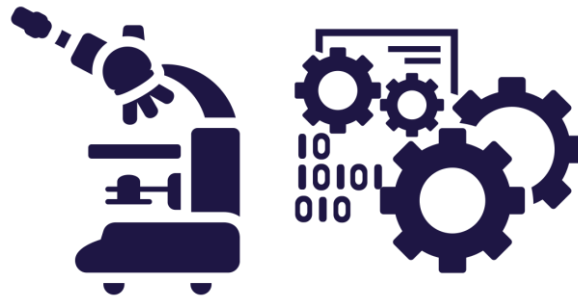


Data, mapa a teritorium



Data Science

Mezioborová technicky orientovaná oblast, která se zabývá inovativním získáváním businessově užitečných znalostí z velmi heterogenních zdrojů dat.



*The key word in “Data Science” is not Data, it is Science.
(Jeff Leek, 2013)*

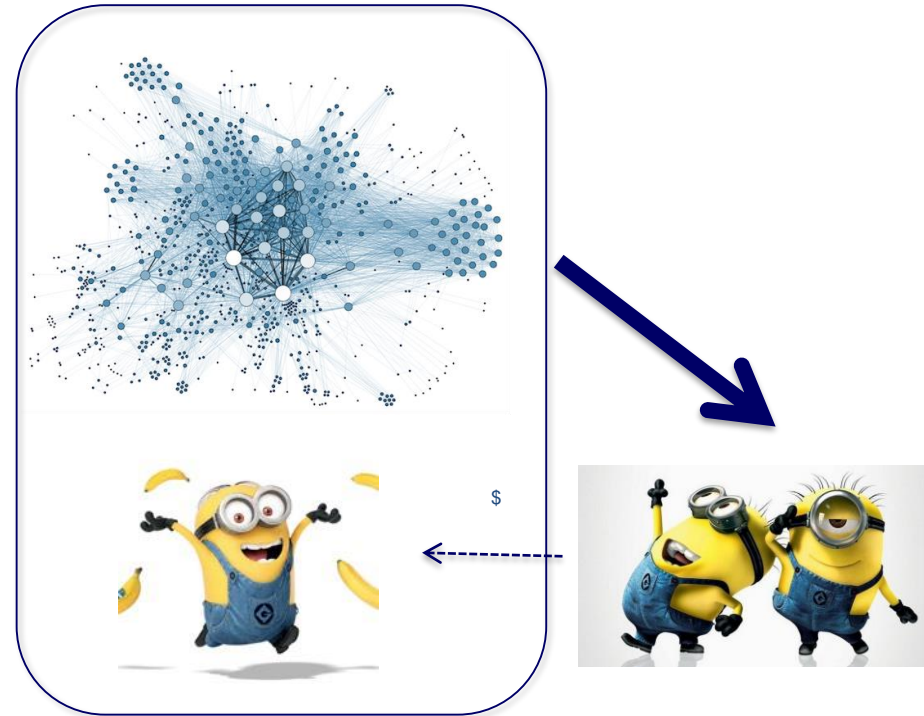
Tři „bigdatí“ módy telco společnosti

Tři „bigdatí“ módy telco společnosti (1)

Mód 1

Prodávání dat

- o telco data má zájem velké množství subjektů
- anonymizovaná telco data jsou prodávána externě
- telco společnost využívá jen okrajově potenciál svých vlastních dat
- je anonymizace u některých dat skutečně bezpečná?

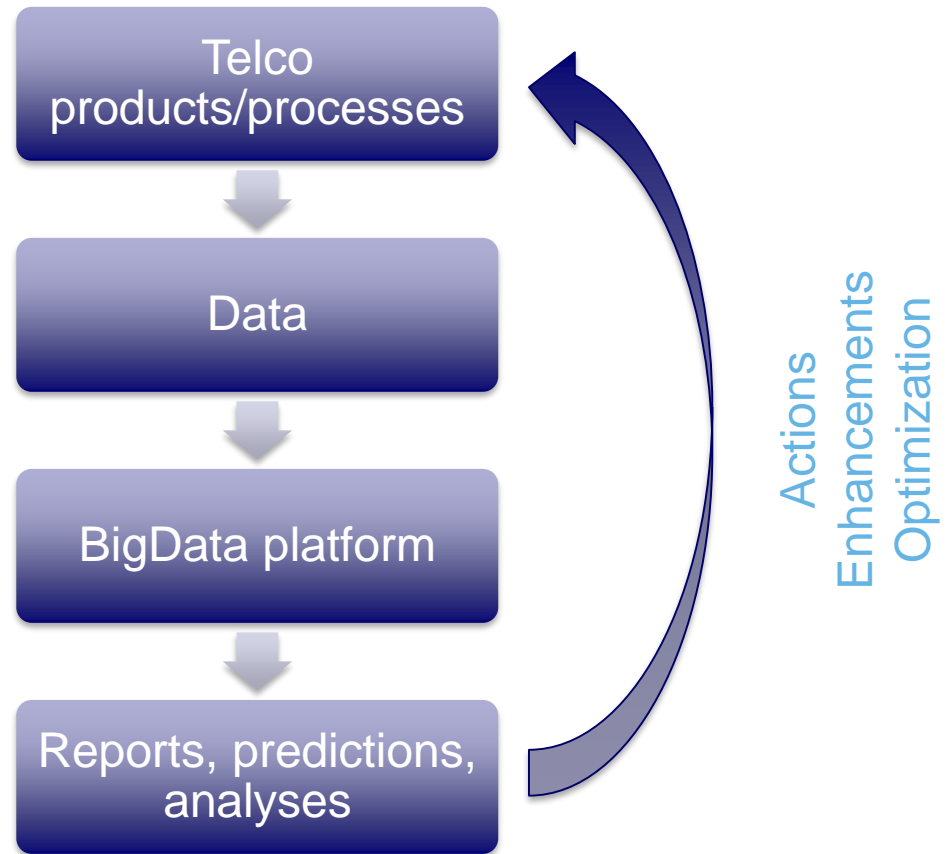


Tři „bigdatí“ módy telco společnosti (2)

Mód 2

Daty řízená optimalizace

- datové produkty vylepšují a optimalizují *interní* procesy a produkty stávajícího businessu
- využití analytické BigData platformy pro BI, plánování, predikci, sledování výkonu, optimalizaci procesů, apod.
- „data science“ může být částečně automatizována různými „out-of-the-box“ řešeními

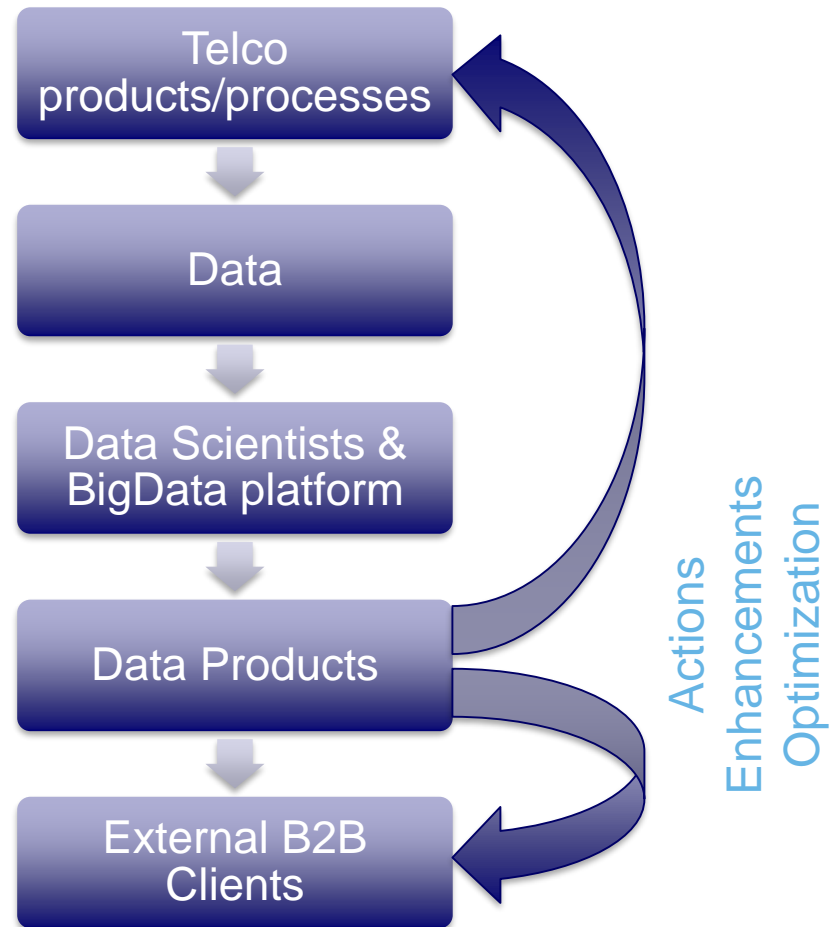


Tři „bigdatí“ módy telco společnosti (3)

Mód 3

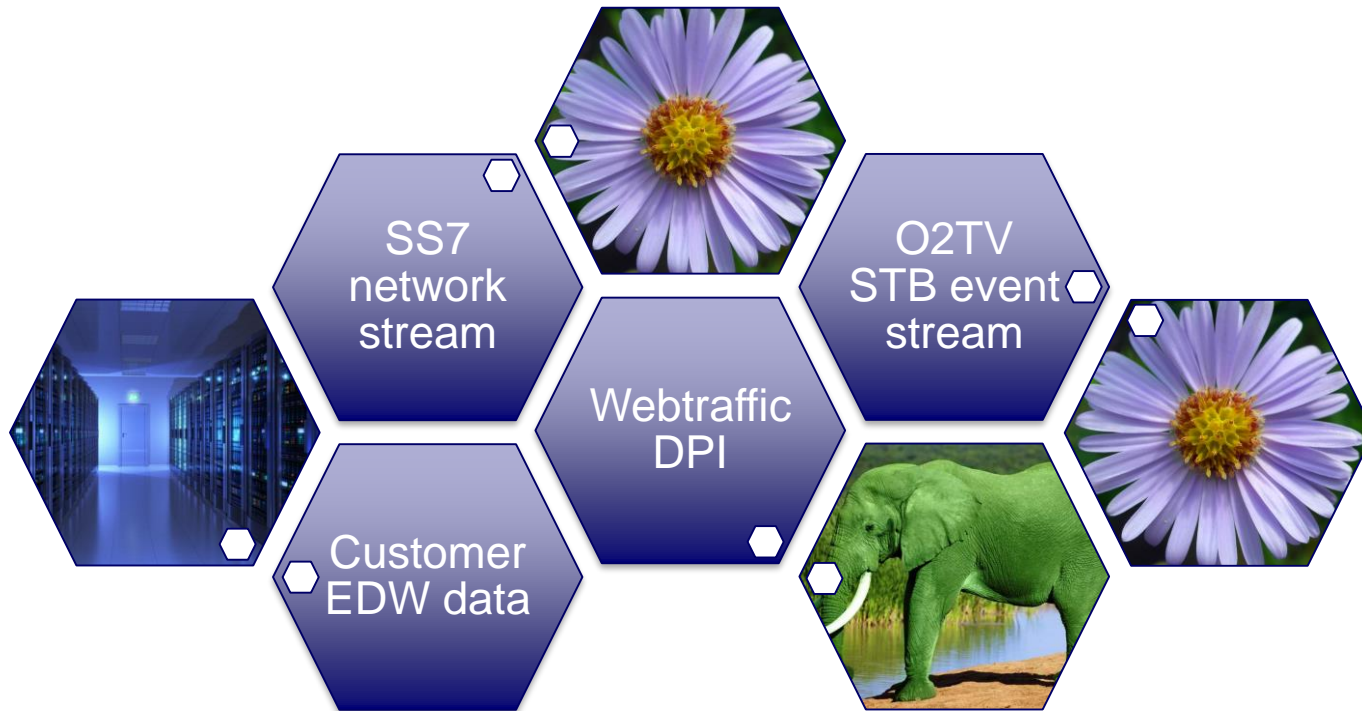
Monetizace datových produktů

- nové zdroje revenue založené na *datových produktech*
- telco společnost *externě* prodává B2B poznatky, vhledy, analýzy či automatizovaná řešení založená na vlastních datech, nikoli data samotná
- spojení BigData platformy a vlastního rozvinutého Data Science týmu má inovační potenciál (podobně jako věda ve smyslu *science* jej má)

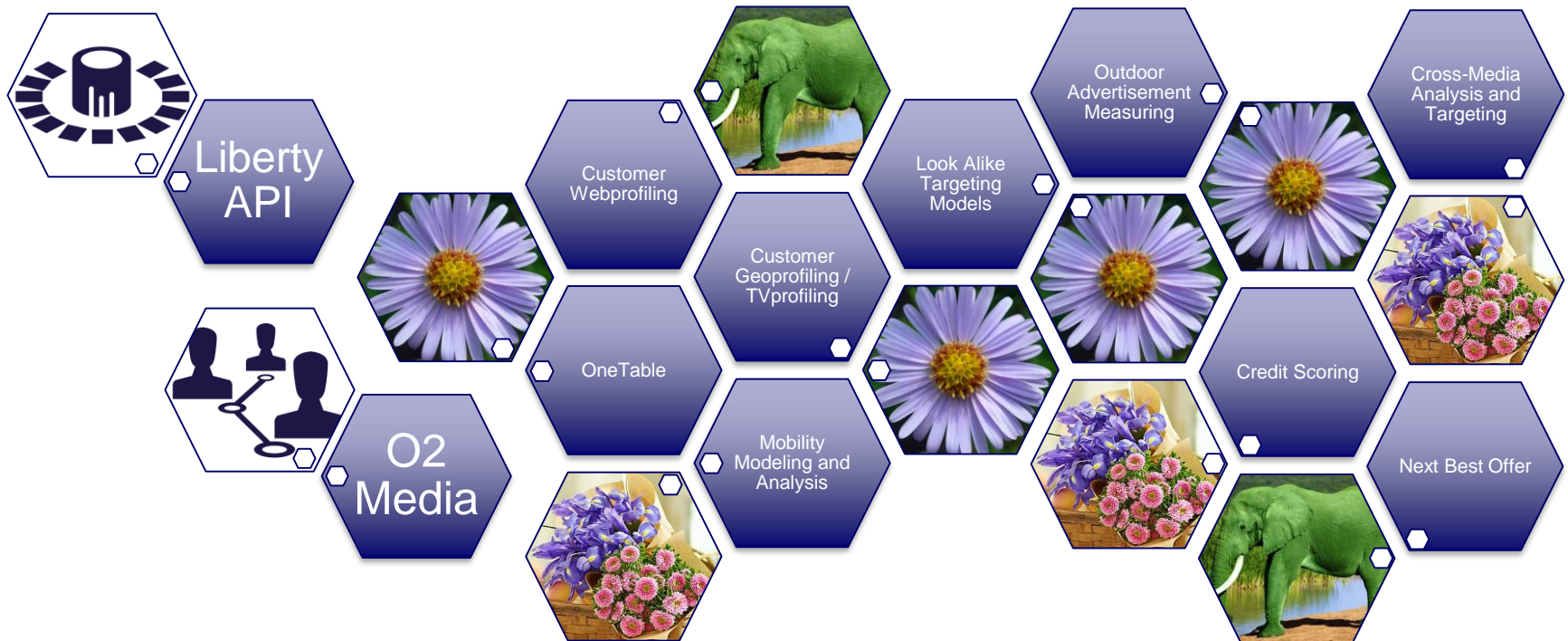


O2 Czech Republic „mód 3“

O2 zdroje Big Dat



Vybrané datové produkty/projekty O2



Profilace zákazníků

- průběžné modelování atributů a segmentů rezidentních zákazníků
- určené jak pro marketingovou segmentaci, cílení, BI analýzy apod., tak jako vstup (prediktory) do většiny dalších datových produktů
- nad všemi datovými zdroji
- celá řada vzájemně provázaných modelů a filtrů různých typů (prediktivních, deskriptivních, statistických, heuristických)
- rychlé/pomalé atributy
- interpretovatelné/latentní atributy
- důraz na automatizovatelnost a agregaci výstupů ve OneTable

Profilace zákazníků – deep learning

- využití deep learning metod k latentní klasifikaci zákaznického chování či dalších jevů
- aplikace metod s komplexními neuronovými sítěmi – např. Restricted Boltzmann Machine nebo Long Short-Term Memory
- použití jako prediktory u jiných datových produktů (zejména risk scoring)

Next Best Offer

- interně vyvinutý recommendation system na produkty pro rezidentní zákaznky
- vlastní algoritmus – kombinace unsupervised clusteringu a supervised metod
- využití všech dostupných prediktorů
- velmi velké maticové operace
- zvýšení ARPU všech kanálů, které NBO používají
- kontinuální sběr zpětné vazby a adaptace

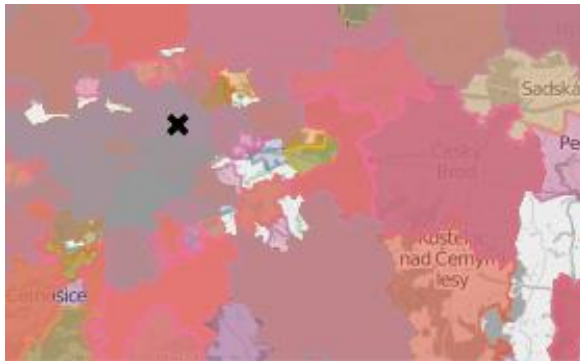
Mobility produkty



měření outdoorové reklamy

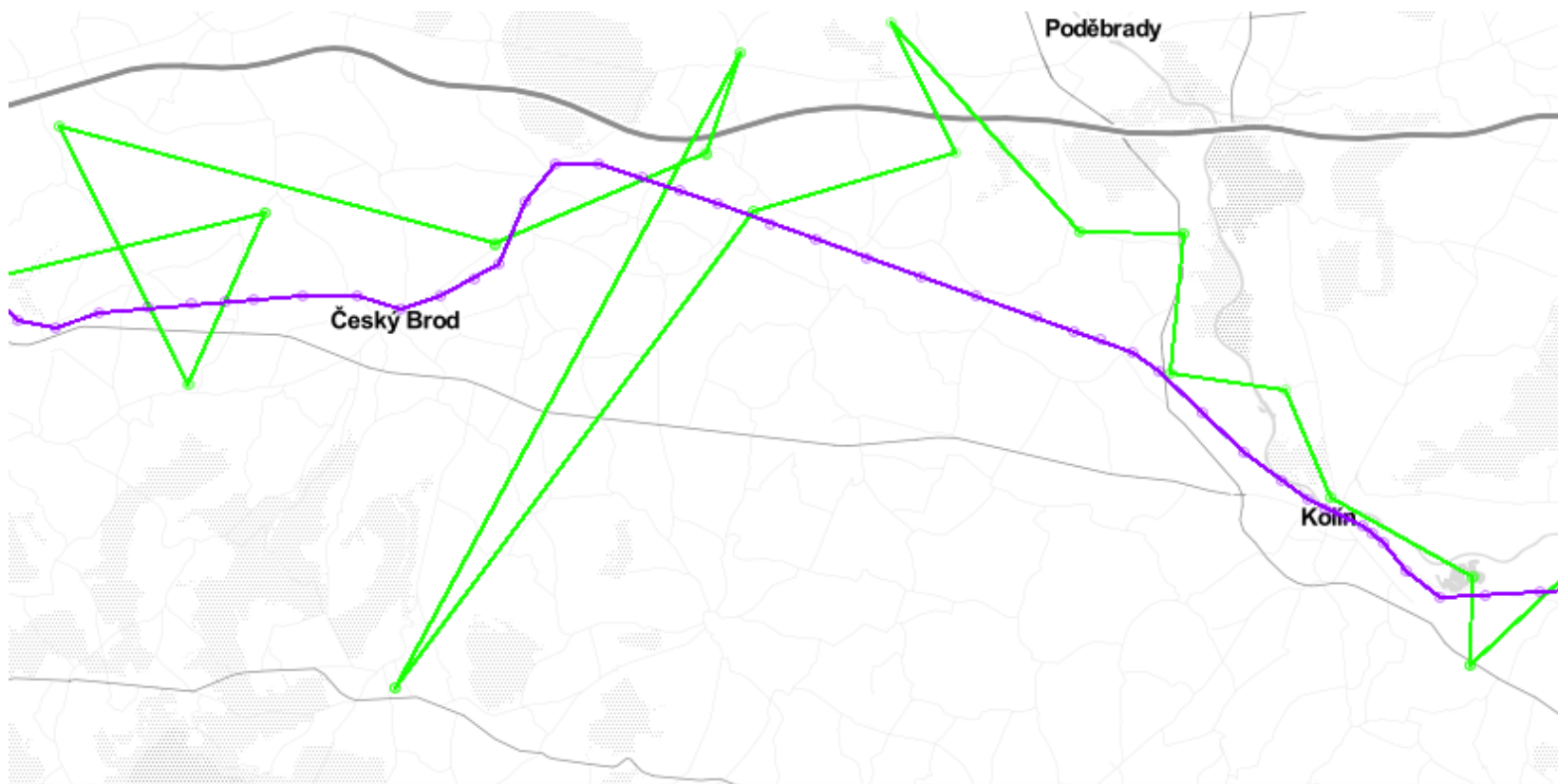


dopravní analýzy



analýzy návštěvnosti, tranzitů, spádovosti, apod.

Mobility produkty



Liberty API – Socio-Demo a Mobility



Počáteční lokalita

Baarova

Cílová lokalita

Bohdalec-jih

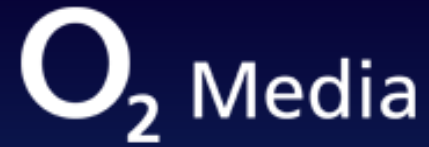
Request

<https://developer.o2.cz/mobility/sandbox/api/transit/127752/132195/?uniques=1>

Response

```
{  
  "count": "2033"  
}
```

Smart Targeting



- supervised a unsupervised machine learning pro cílení B2B kampaní (SMS, online)
- predikce prokliku reklamy na základě všech dostupných prediktorů
- učení na základě zpětné vazby

SNA a detekce komunit z CDR

- prediktivní model pro detekci komunit (rodiny, zájmové skupiny, apod.) a Social Network Analysis
- marketingové využití
- vstup do dalších datových produktů
- aplikace metod používaných v oblasti Natural Language Processing na grafovou analýzu záznamů o volání
- využití neuronových sítí a grafově-analytických nástrojů

Webprofiling zájmů

- segmentace podle zájmového chování zákazníků na webu
- automatický webscraping a strojová klasifikace webových stránek do zájmových kategorií na základě zpracování textu
- “starý dobrý bag-of-words & SVM“ přístup s využitím n-gramů a TF/IDF – sice dávno překonaný, ale pro tyto účely postačující
- statistické modelování zájmu zákazníků podle zájmových kategorií navštívených webů

Pokročilé zpracování textu

- NLP zpracování textových dat např. z reklamačních protokolů či ASR přepisů hovorů z call center
- využití pro BI a jako vstup do dalších datových produktů (např. NBO)
- základní modely pro statistický topic modelling – Latent Dirichlet Allocation
- pokročilé metody jazykového modelování: word2vec (skip-gramy), doc2vec, fastText, lda2vec

Děkuji za pozornost

**Connect with me @JanRomportl
jan.romportl@o2.cz**

