



Dopravní nehody v ČR

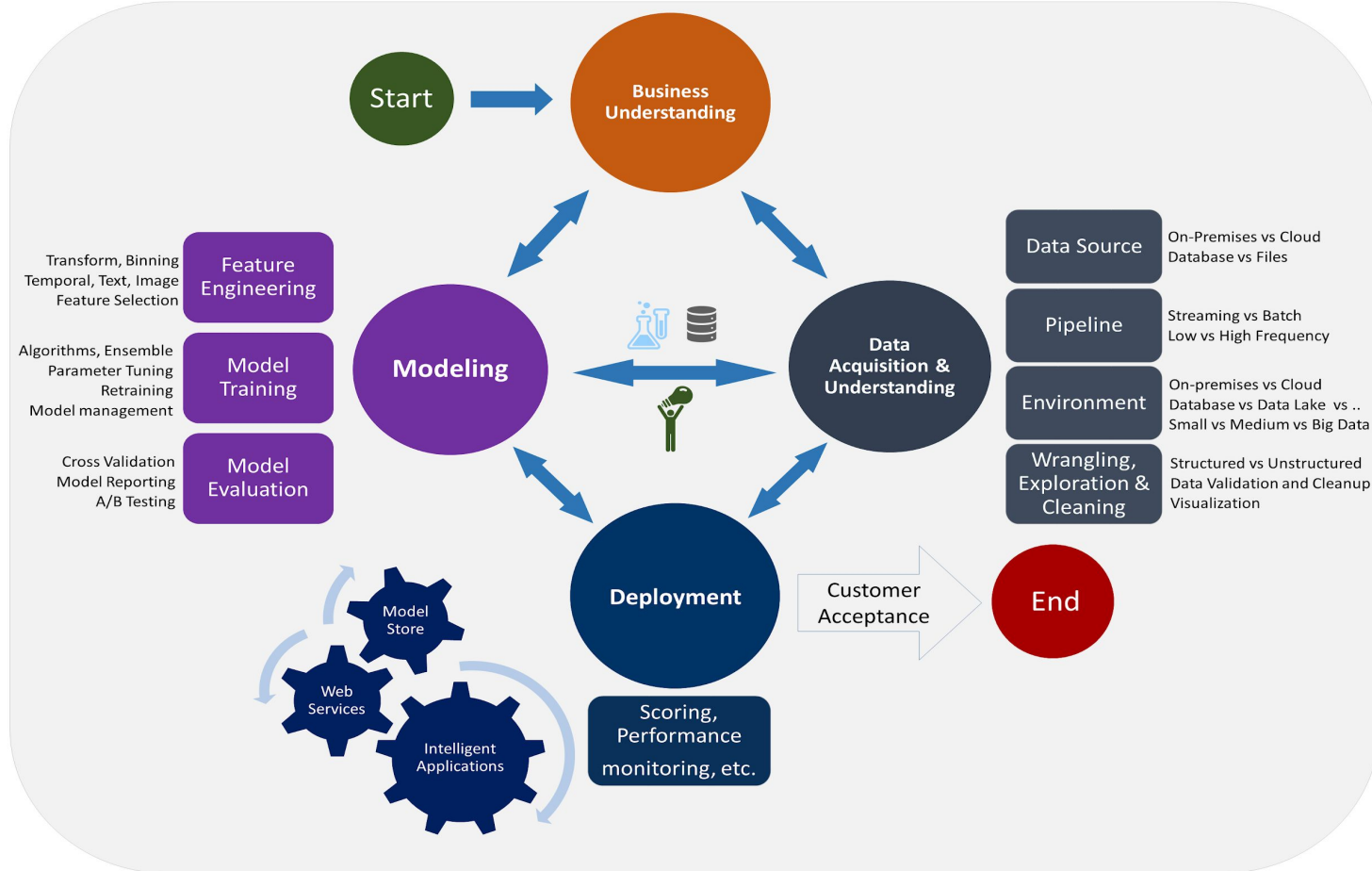
Tým

- Bozděchová Slávka
- Jozífek Michal
- Kovalčuk Jiří
- Martínek Tomáš
- Žydyk Roman

Agenda

- Náš postup
- Business understanding
- Data exploration
- Data preparation
- Modelling
- Evaluation

Data Science Lifecycle



Náš postup

1. Stažení dat získaných Policií ČR
2. Přejmenování a sjednocení do jedné složky
3. Sjednocení dle měsíce, města (s přidáním headeru)
4. Raw data exploration
5. Mapování + vyčištění
6. Explorace namapovaných dat
7. Bucket + indexer + one/hot (analytická tabulka)
8. Modelování
9. Vyhodnocení

1

Business understanding

1 Business understanding

- Výběr datasetu
 - Česká data
 - Málo užitečných open dat




Data o dopravních nehodách v ČR za rok 2019

2 - Data understanding and acquisition

- Policie ČR
- Každý měsíc, každý kraj - posledních 12 let
- Sloučeno 168 CSV souborů v Pythonu (~500k řádků)
- **Label** = Charakter nehody
- **Cíl modelu** - určit atributy a jejich konkrétní dopad vedoucí k nehodě s úmrtím/i
- **Využitelnost** - prevence výskytu takových nehod

2

Explorace dat

- 
- vyber dat, zisk dat, forma dat, prvotni uprava dat (mnoho csv), prvotni explorace, zjistení které sloupce potreba (vse na malých vzorcích)
 - druhotna uprava dat - mapovani hodnot, nully, sloupce, pocty unikatnich hodnot, datove typy, imputovani, dropovani
 - explorace upravenych dat - pocty radku, sloupcu, datove typy, statistika, histogramy, korelace, testovani bucketovani + vytvoreni pipeline a analyticke tabulky
 - modelovani - subsety, vybrane metody, trenovani, predikce, vypsani koeficientu a stromu
 - evaluace modelu - vysledky, nejlepsi model, jejich parametry, cross validate - interpretace

2:22 PM

Děkujeme.

Prostor pro Vaše dotazy