

O₂

Prezentácia projektu

Daniel Kurucz

Zadanie

Úloha:

1. Prosim stiahnite si ulohu z Kaggle <https://www.kaggle.com/yelp-dataset/yelp-dataset>.
2. Vyberte si programovací jazyk (ideálne Python alebo R).
3. Ukazte prosim vase programovacie zrucnosti a pracu s datovymi vstupmi a popiste co zaujimaveho vidite v datach.
4. Zamyslite sa nad business vyuzitim
5. Aplikujte machine learning algoritmus, popiste a vysvetlite problem ktory chcete vysvetlit
6. Prosim zaslite nam kod v Jupyter Notebooku alebo VCS (GitHub, GitLab, etc.), aby sme vedeli zreprodukovat vase riesenie
7. Po zhodnoteni ulohy vas budeme kontaktovat a prejdeme si spolocne vysledky

[extra uloha - volitelna]

- Vytvorte geo-priestorovu vizualizaciu
- Provnajnie viac ako jedoho modelu strojoveho ucenia (machine learning)

Práca s dátami za účelom vytvorenia kampaní

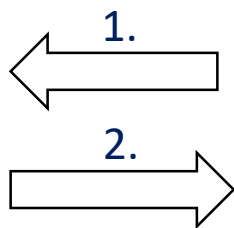
Kampaň číslo 1. – mobilný paušál O2 dáta

Zefektívnenie predaja produktu a uvedenie do pozornosti širšieho spektra zákazníkov kt. „dátujú“. Zameranie sa na top kategórie (podľa počtu hodnotení) v ktorých boli napísané reviews. Časovo by kampaň bola smerovaná podľa priemerného času napísania review v spomínanej kategórii. Identifikovanie O2 zákazníkov v daných kategóriách by sa dialo na základe polohy SIM karty

Kampaň číslo 2. – Vianočná kampaň

Zvýšenie predaju zvýhodnených produktov O2 v čase Vianoc. Vyselektovanie zákazníkov v mesiacoch 11,12 kt. napísali review v kategóriách ktoré sa svojimi produktami podobajú ponuke O2

	state	categories	business_id	Avg_hour
319	AB	Restaurants	524	12.318702
152	AB	Food	236	14.868644
331	AB	Shopping	198	16.242424
908	AZ	Restaurants	503	12.582505
930	AZ	Shopping	315	15.698413
633	AZ	Food	267	14.022472
1443	CA	Restaurants	208	12.716346
1461	CA	Shopping	127	15.834646
1228	CA	Food	123	14.447154
1776	DE	Restaurants	226	15.769912



	review_id	user_id	business_id	date	categories
313	jC-fGfx-YLqxVBcyTAd4Pw	EBa-0-6AKoy6jziNexDJtg	W4ZEKkva9HpAdZG88juwyQ	12	Bars
314	CzvOSoP8WTJgtkqyT4i7zA	nFiYurEbPZBAOCURMrBgag	W4ZEKkva9HpAdZG88juwyQ	12	Bars
315	DwQeq3W2p0QbAVSubfg58A	BeS9uhpmc6lcap_Eso9Lqw	W4ZEKkva9HpAdZG88juwyQ	12	Bars
316	O9Ky-Cn-sGoD98ETyOb2zg	LQ2CDCzIkXrUMXnh-W9QSg	W4ZEKkva9HpAdZG88juwyQ	11	Bars
317	LKH11_albFux0Uib2CLlhQ	CcXSCGzlhN8Of6xwafPklQ	W4ZEKkva9HpAdZG88juwyQ	11	Bars

Práca s dátami za účelom vytvorenia kampaní

3. Podpora aktívnych kampaní na základe kľúčových slov

Výber zákazníkov na základe kľúčových v prislúchajúcich reviews / tips ktoré boli napísané. Je potrebné očistenie reviews o bežné slová ako AND, I, GOOD atď...

4. Oslovenie influencerov

Oslovenie influencerov za účelom spolupráce. Sú to ľudia kt. sú active v počte reviews / majú veľkú fanúšikovskú základňu

5. O2 pods kampaň

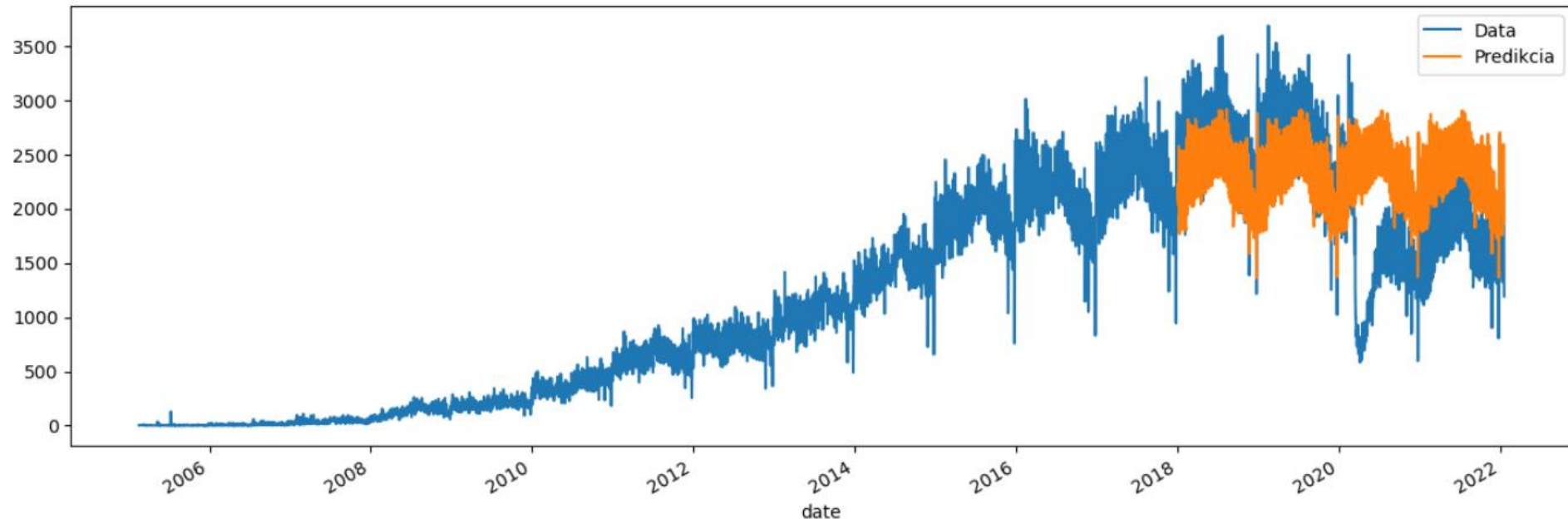
„Výber telefónov
bol na v tomto
obchode veľmi
malý.“



Porovnanie predikčných modelov

Predikčný model za použitia XGboost za účelom predikovania počtov reviews

1. Očistenie df
2. Rozdelenie df na train a test vzorku ($< 1.1.2018$, $\geq 1.1.2018$)
3. Priradenie tzv. „features“ na preskúvanie df dát podľa hodiny, dňa, mesiaca...
4. Nastavenie modelu, ($n_estimators = 1000$ (počet DT podľa ktorých sa bude model učiť a zlepšovať), $early_stopping_rounds = 50$ (počet iterácií na základe ktorých sa bude diať variabilita skúmaných hodnôt), $learning_rate = 0.01$ (rýchlosť učenia modelu))
5. Uplatnenie predikčnej funkcie `test['prediction'] = reg.predict(X_test)`



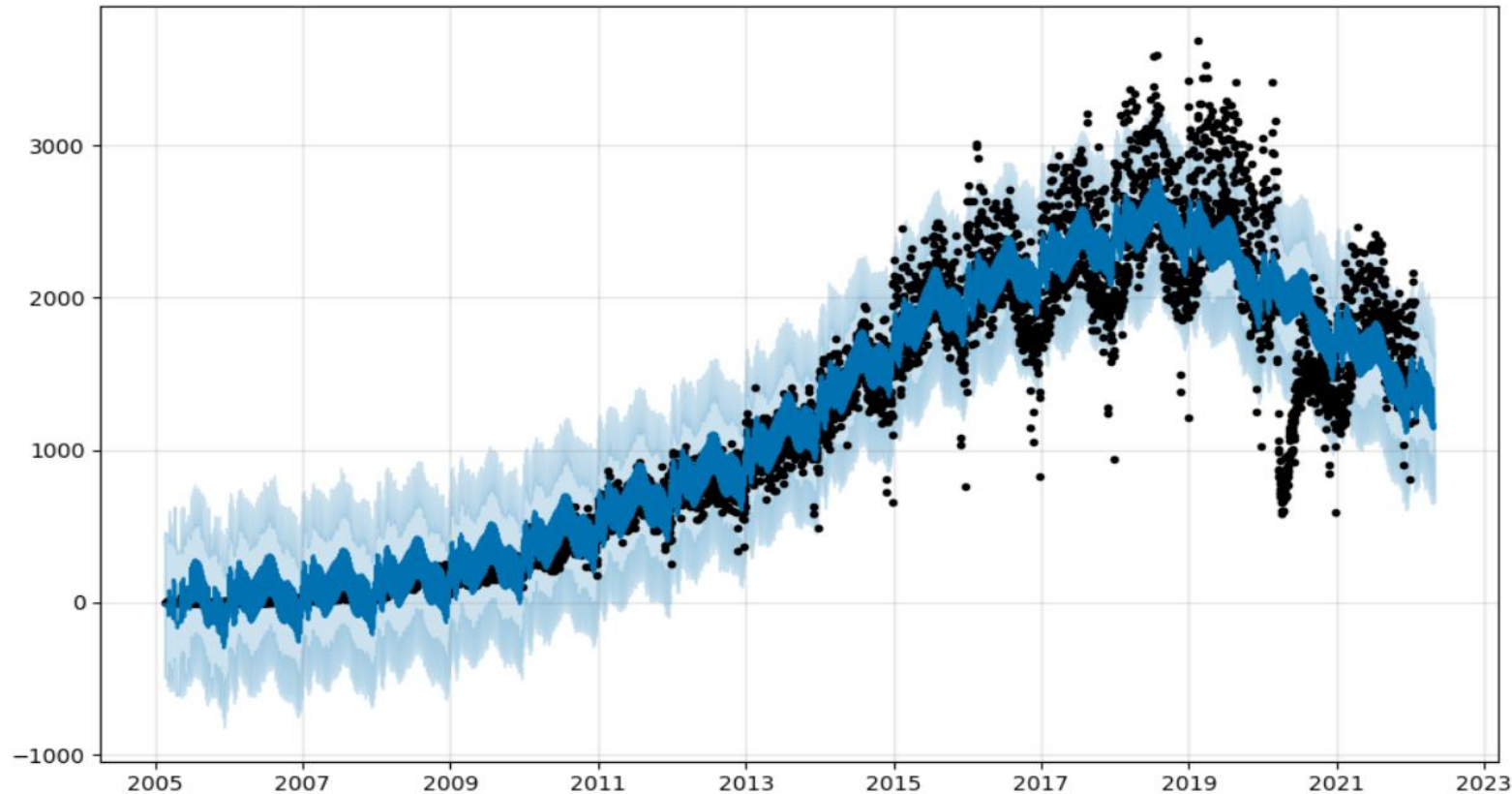
Porovnanie predikčných modelov

Predikčný model za použitia knižnice Prophet

1. Očistenie df
2. Aplikácia do modelu

Prophet(interval_width=0.95 (orezáva horný a spodný interval v pozadí), daily_seasonality=True (pozerá sa na denný vývoj resp. trend))

```
future = m.make_future_dataframe(periods=100,freq='D,')
```



Geo mapa

Geo zobrazenie zamerané na počty hodnotení podľa zemepisnej šírky a dĺžky

