

Prezentácia projektu



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

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 Ear pods kampaň

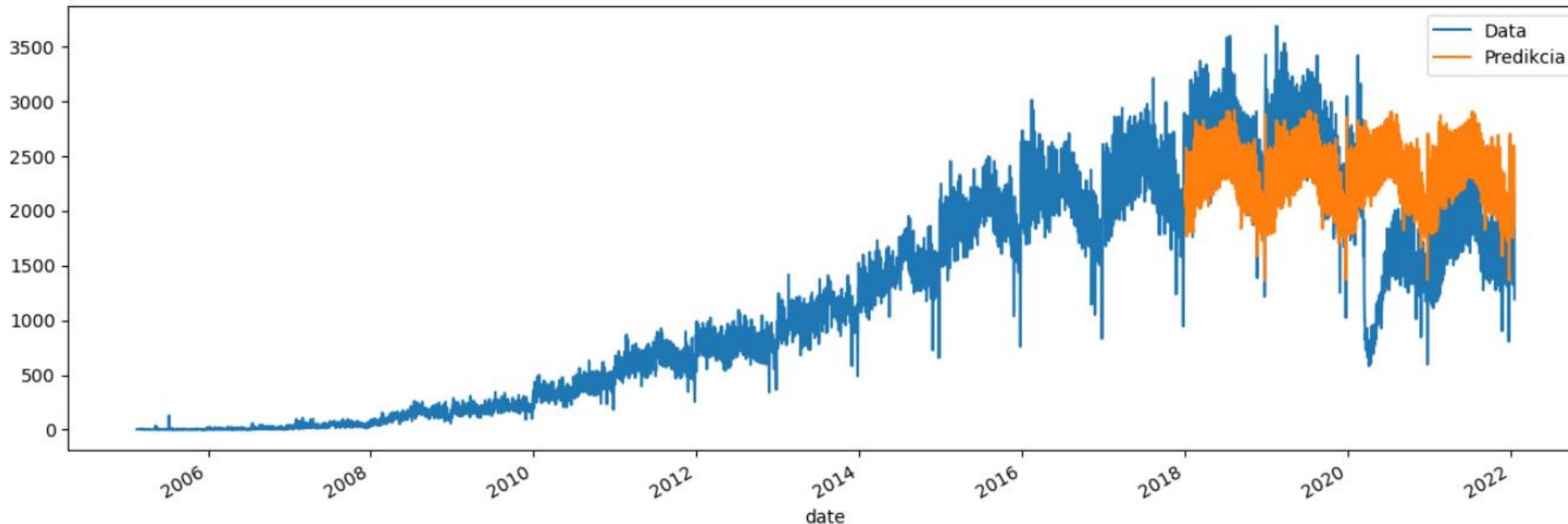
Vyselektovanie kategórie šport / inej v ktorej by sa mohli používať Ear pods. Na základe avg času review smerovať na tieto lokality geo lokačnú kampaň / na ľudí kt. napísali review



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 preskúmanie 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)`



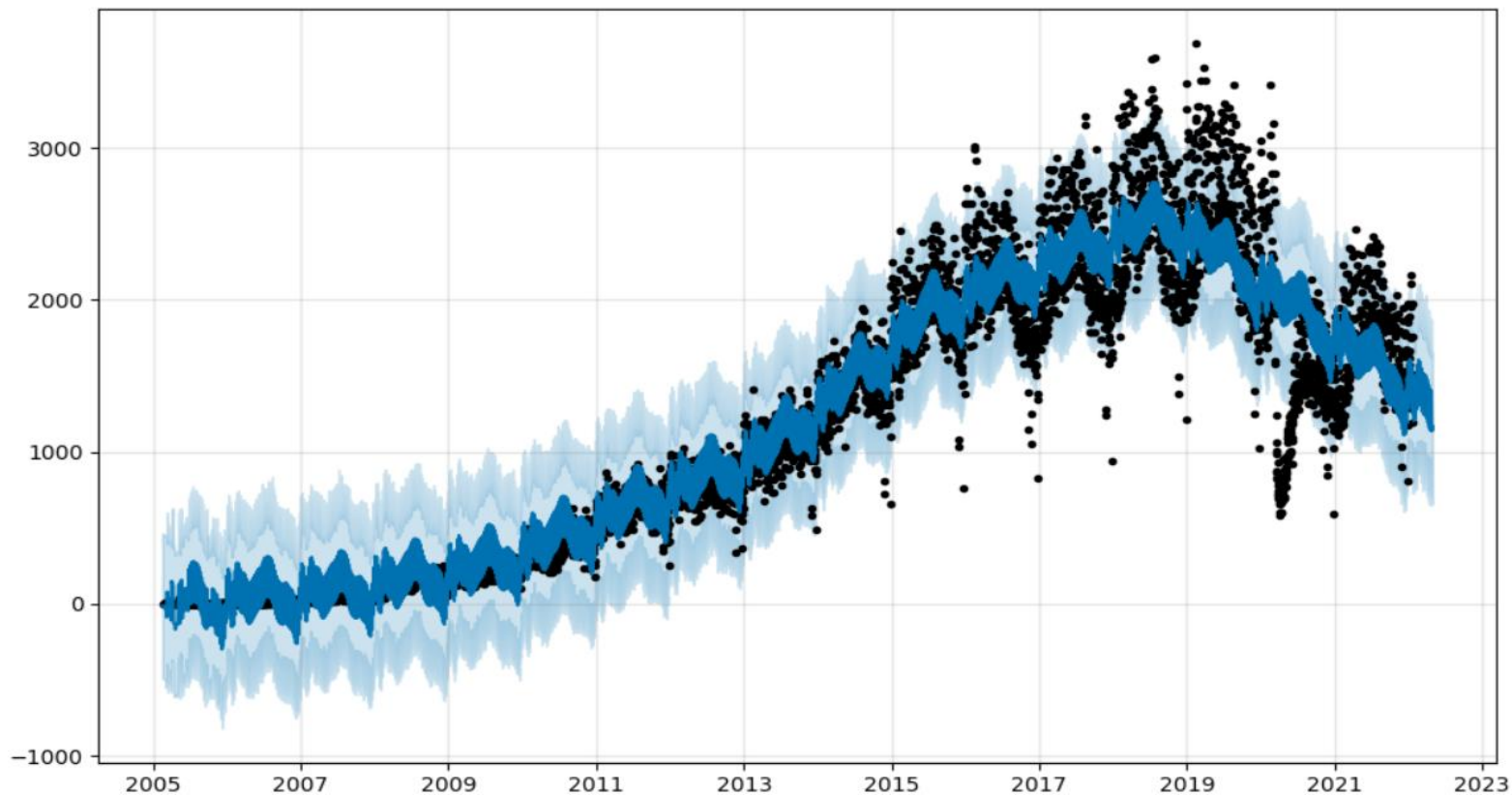
Porovnanie predikčných modelov

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

