

Predikcia nákladov zdravotnej starostlivosti pomocou vybraných metód strojového učenia

Študent: Bc. Pavel Lukačik

Vedúci: RNDr. Ľubomír Antoni, PhD.

Motivácia

Ciele

1. Vytvoriť prehľad aktuálne existujúcich metód strojového učenia na predikciu nákladov zdravotnej starostlivosti.
2. Navrhnuť a implementovať kombináciu metód strojového učenia vhodnú na riešenie predikcie nákladov zdravotnej starostlivosti.
3. Porovnať presnosť predikcie navrhnutého riešenia s inými dostupnými štúdiami.

Prehľad metód

- BERTSIMAS, D., BJARNADOTTIR, M. V., KANE, M. A. a kol. (2008). Algorithmic prediction of health-care costs. *Operations Research*, 56(6), 1382-1392.
- MORID, M.A, KAWAMATO, K. a kol. (2017). Supervised Learning Methods for Predicting Healthcare Costs: Systematic Literature Review and Empirical Evaluation. *AMIA Annu Symp Proc*. 2017; 2017: 1312–1321.

Výber metód

Baseline 1 - suma za posledný rok

Baseline 2 - priemer súm za posledné tri roky

Neurónová sieť - rôzne typy a architektúry

Gradient boosting

Porovnanie presnosti

- Mean absolute percentage error (MAPE)
- R^2 - Pearsonová korelácia medzi skutočnými a predikovanými hodnotami
- Hit Ratio
- The Absolute Prediction Error

Dataset (základná verzia)

- vzorka 17 000 poistencov súkromnej zdravotnej poisťovne
- obdobie - január 2010 až december 2013
- mesačné sumy nákladov za lieky vo všetkých kategóriách ATC
- obsahuje len pacientov s plnou históriou v danom období

Nástroje

- Python
- XGBoost
- Keras



dmlc
XGBoost



Keras

Najbližšie úlohy

- naprogramovať a vyhodnotiť BASELINE1 a BASELINE2
- naprogramovať doprednú neurónovú sieť a vyhodnotiť
- porozmýšľať nad iným modelom neurónovej siete
- naštudovať knižnicu XGBoost

Ďakujem za pozornosť
