

# Predikcia nákladov zdravotnej starostlivosti pomocou vybraných metód strojového učenia

---

Študent: Bc. Pavel Lukačik

Vedúci: RNDr. Ľubomír Antoni, PhD.

# Motivácia

- Rastúce náklady na zdravotnú starostlivosť predstavujú jednu z aktuálnych výziev v celosvetovom meradle
- Jedným z možných nástrojov na kontrolu nákladov na zdravotnú starostlivosť je presná predikcia pravdepodobných budúcich nákladov jednotlivcov
- Prediktívne modelovanie môže zvýšiť účinnosť systému zdravotnej starostlivosti a umožniť efektívnejšiu alokáciu zdrojov

# 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.

# Štúdium literatúry

- metódy učenia s učiteľom
- atribúty nákladové (+ nenákladové)
- medicínske údaje zvyčajne vykazujú vysoký stupeň odchýlok v rámci jednotlivých pacientov.
- príslušné pravdepodobnostné rozdelenie údajov nie je normálne
- najúspešnejšie Gradient Boosting, Umelá neurónová sieť
- návrhy: použiť hĺbkové učenie, štrukturálnu analýzu, pridanie nenákladových atribútov

# Výber metód

Baseline 1 - suma za posledný rok

Baseline 2 - priemer súm za posledné tri roky

Neurónová sieť - dopredná

Neurónová sieť - CNN (+ Inception-Res)

Extreme Gradient boosting

Light Gradient boosting

	Actual Bucket				
Predicted Bucket	0	2	4	6	8
	1	0	2	4	6
	2	1	0	2	4
	3	2	1	0	2
	4	3	2	1	0

$$R^2 = 1 - \frac{\sum_i (a_i - p_i)^2}{\sum_i (a_i - \bar{a})^2}$$

$$\text{Hit Ratio} = \frac{\text{Number of members with correct predicted bucket}}{\text{Total number of members}}$$

$$\text{MAPE} = \frac{\frac{\sum_i |a_i - p_i|}{n}}{\bar{a}}$$

## Hodnotenie úspešnosti

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

$$\text{MAE} = \frac{\sum_i |a_i - p_i|}{n}$$

# 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
- LGBM
- Keras





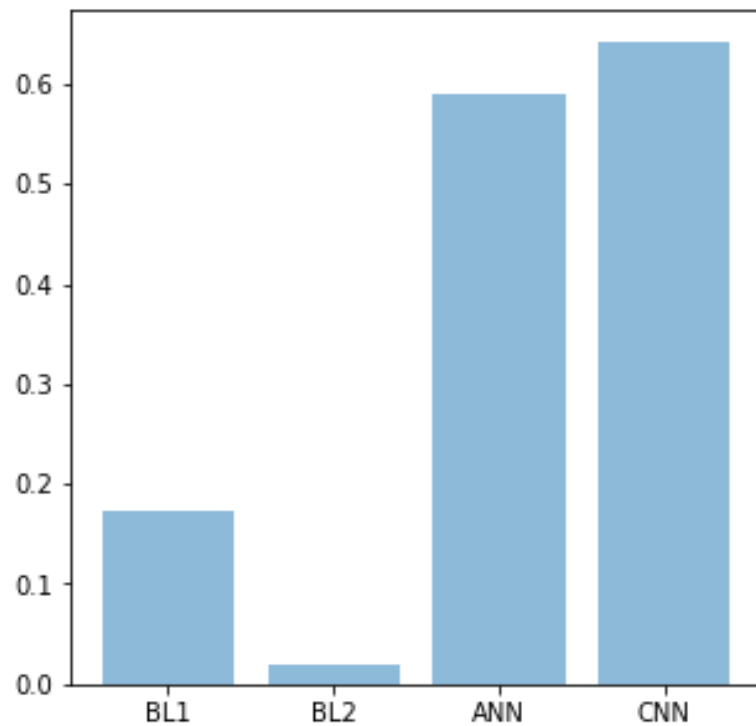
# Súčasný stav

- spracovaný prehľad literatúry zameranej na predikciu nákladov zdravotnej starostlivosti
- implementované riešenia BL1 a BL2
- implementované riešenia pomocou doprednej a konvolučnej neurónovej siete
- optimalizácia parametrov neurónových sietí
- prebieha: štúdium metód XGBoost, LGBM a ich implementácia

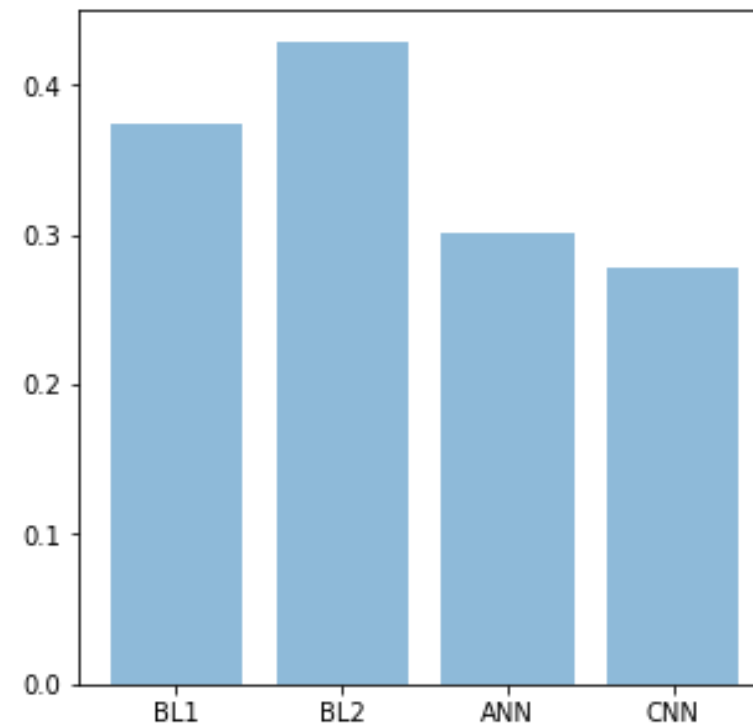
# Výsledky (orientačné)

---

R<sup>2</sup> score



MAPE



Ďakujem za pozornosť

---