

# Predikcia budúcich nákladov na pacienta

Marián Kravec  
školiteľ: MSc. František Dráček

# Obsah

## Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich záznamov

Predikcia budúcich záznamov

Predikcia cenovej kategórie záznamu

Trénovanie, testovanie a validácia

Trénovanie a testovanie

Validácia

Ďalšie kroky

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich  
záznamov

Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie

Validácia

Ďalšie kroky

# Cieľ práce

Predikovanie budúcej nákladovej skupiny do ktorej bude pacient v budúcnosti patriť na základe jeho minulosti a súčasného zdravotného stavu

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony  
Diagnózy  
Lieky

Predikcia nákladov

Proces predikcie  
Výpočet počtu budúcich  
záznamov  
Predikcia budúcich  
záznamov  
Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

# Obsah

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich záznamov

Predikcia budúcich záznamov

Predikcia cenovej kategórie záznamu

Trénovanie, testovanie a validácia

Trénovanie a testovanie

Validácia

Ďalšie kroky

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich  
záznamov

Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

# Dáta

- ▶ Zdroj dát: NCZI
- ▶ Typ dát: poistné dáta
- ▶ Jeden záznam: informácia o jednom výkone/lieku konkrétneho pacienta
- ▶ Datasets:
  - ▶ Dataset ambulantnej zdravotnej starostlivosti
  - ▶ Dataset ústavnej zdravotnej starostlivosti
  - ▶ Dataset predpísaných liekov
- ▶ V súčasnosti využívaná malá anonymizovaná vzorka

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony  
Diagnózy  
Lieky

Predikcia nákladov

Proces predikcie  
Výpočet počtu budúcich  
záznamov  
Predikcia budúcich  
záznamov  
Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

# Najdôležitejšie údaje pre časový rad

- ▶ Čas od prvého výkonu
- ▶ Vek pacienta
- ▶ Medicínsky výkon
- ▶ Diagnóza
- ▶ Liek
- ▶ Cena výkonu/lieku

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony  
Diagnózy  
Lieky

Predikcia nákladov

Proces predikcie  
Výpočet počtu budúcich  
záznamov  
Predikcia budúcich  
záznamov  
Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

# Obsah

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich záznamov

Predikcia budúcich záznamov

Predikcia cenovej kategórie záznamu

Trénovanie, testovanie a validácia

Trénovanie a testovanie

Validácia

Ďalšie kroky

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich  
záznamov

Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

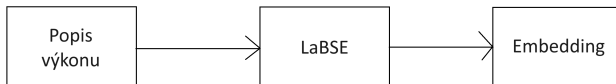
Trénovanie a testovanie

Validácia

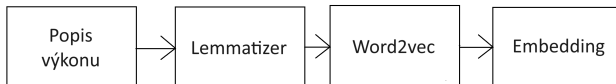
Ďalšie kroky

# Výkony

- Language-agnostic BERT sentence embedding model (LaBSE)



- Lemmatizer + Word2vec model



Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

**Výkony**

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich  
záznamov

Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

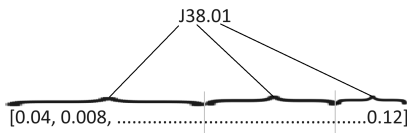


# Diagnózy

## ► ICD-10-CM kódy

[znak][dvojčíslenie].[dvojčíslenie]

- J - Choroby dýchacej sústavy
- J3 - Iné choroby dýchacích ciest
- J38 - Choroba hlasiviek a hrtana
- J38.0 - Obrna hlasiviek a hrtana
- J38.01 - Jednostranná čiastočná obrna hlasiviek a hrtana



Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Átá

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich  
záznamov

Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

## ► ATC kód

[znak][dvojčísle][dvojica znakov][dvojčísle]

- N - Centrálna nervová sústava
- N02 - Analgetiká
- N02B - Iné analgetiká a antipyretiká
- N02BA - Kyselina salicylová a deriváty
- N02BA01 - Acylpyrín

# Obsah

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

**Predikcia nákladov**

Proces predikcie

Výpočet počtu budúcich záznamov

Predikcia budúcich záznamov

Predikcia cenovej kategórie záznamu

**Trénovanie, testovanie a validácia**

Trénovanie a testovanie

Validácia

**Ďalšie kroky**

**Predikcia budúcich  
nákladov na  
pacienta**

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

**Predikcia nákladov**

Proces predikcie

Výpočet počtu budúcich  
záznamov

Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

**Trénovanie,  
testovanie a  
validácia**

Trénovanie a testovanie

Validácia

**Ďalšie kroky**

# Proces predikcie

1. načítanie dát pacienta, embedding a normalizácia
2. výpočet počtu záznamov na budúci rok
3. predikcia budúcich záznamov
4. predikcia cenových kategórii budúcich záznamov
5. výpočet celkovej cenovej kategórie

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony  
Diagnózy  
Lieky

Predikcia nákladov

**Proces predikcie**

Výpočet počtu budúcich  
záznamov

Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

# Výpočet počtu budúcich záznamov

- ▶ Možnosť A) lineárna regresia počtu záznamov z predchádzajúcich rokov
- ▶ Možnosť B) generovanie nových záznamov kým generovaný timestamp nepresiahne rok od posledného skutočného záznamu

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich  
záznamov

Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

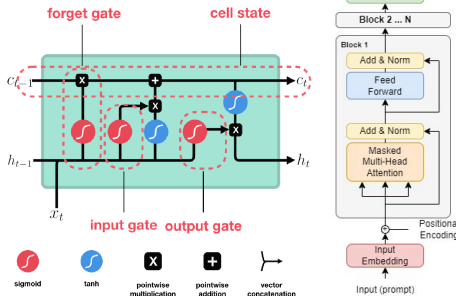
Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

# Predikcia budúcich záznamov

- ▶ využitie fixného maximálneho okna pre umožnenie paralelizácie trénovania
- ▶ testované modely
  - ▶ Long short-term memory
  - ▶ Decoder-only transformer



Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich  
záznamov

Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

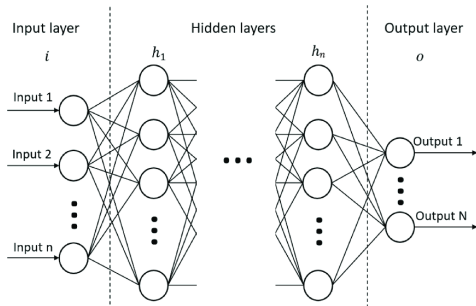
Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

# Predikcia cenovej kategórie záznamu

- ▶ namiesto konkrétnej ceny predikujeme cenovú kategóriu
- ▶ použitý model: Multilayer perceptron



Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony  
Diagnózy  
Lieky

Predikcia nákladov

Proces predikcie  
Výpočet počtu budúcich  
záznamov  
Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

# Obsah

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich záznamov

Predikcia budúcich záznamov

Predikcia cenovej kategórie záznamu

Trénovanie, testovanie a validácia

Trénovanie a testovanie

Validácia

Ďalšie kroky

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich  
záznamov

Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie

Validácia

Ďalšie kroky



# Trénovanie a testovanie

- ▶ predikčné modely sa trénujú a testujú samostatne
- ▶ ako chyba predikcie záznamu sa využíva váhovaná euklidovská vzdialenosť od reálneho záznamu
- ▶ ako chyba predikcie cenovej kategórie sa využíva euklidovská vzdialenosť od one-hot vektora skutočnej cenovej kategórie

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony  
Diagnózy  
Lieky

Predikcia nákladov

Proces predikcie  
Výpočet počtu budúcich  
záznamov  
Predikcia budúcich  
záznamov  
Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

# Validácia

1. rozdelenie záznamov pacienta na posledný rok a staršie
2. výpočet skutočnej celkovej cenovej kategórie pacienta v poslednom roku
3. predikcia počtu záznamov, jednotlivých záznamov, ich cenovej kategórie a celkovej cenovej kategórie
4. výpočet chyby ako euklidovská vzdialenosť medzi predikovanou a skutočnou cenovou kategóriou v poslednom roku

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Údacia

Embedding

Výkony  
Diagnózy  
Lieky

Predikcia nákladov

Proces predikcie  
Výpočet počtu budúcich  
záznamov  
Predikcia budúcich  
záznamov  
Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky

# Obsah

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich záznamov

Predikcia budúcich záznamov

Predikcia cenovej kategórie záznamu

Trénovanie, testovanie a validácia

Trénovanie a testovanie

Validácia

Ďalšie kroky

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony

Diagnózy

Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich  
záznamov

Predikcia budúcich  
záznamov

Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie

Validácia

Ďalšie kroky

# Ďalšie kroky

- ▶ trénovanie, testovanie a nastavovanie hyperparametrov modelov na väčšom datasete serveri
- ▶ výber najlepších hyperparametrov
- ▶ spísanie výsledkov

Predikcia budúcich  
nákladov na  
pacienta

Cieľ práce

Dáta

Embedding

Výkony  
Diagnózy  
Lieky

Predikcia nákladov

Proces predikcie  
Výpočet počtu budúcich  
záznamov  
Predikcia budúcich  
záznamov  
Predikcia cenovej kategórie  
záznamu

Trénovanie,  
testovanie a  
validácia

Trénovanie a testovanie  
Validácia

Ďalšie kroky