

Predikcia budúcich nákladov na pacienta

Marián Kravec
školiteľ: MSc. František Dráček

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
Trénovanie modelu na
predikciu budúcich
záznamov
Trénovanie modelu na
predikciu cenovej kategórie
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
Trénovanie modelu na predikciu budúcich záznamov
Trénovanie modelu na predikciu cenovej kategórie
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
Trénovanie modelu na
predikciu budúcich
záznamov
Trénovanie modelu na
predikciu cenovej kategórie
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
Trénovanie modelu na
predikciu budúcich
záznamov
Trénovanie modelu na
predikciu cenovej kategórie
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
Trénovanie modelu na predikciu budúcich záznamov
Trénovanie modelu na predikciu cenovej kategórie
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
Trénovanie modelu na
predikciu budúcich
záznamov
Trénovanie modelu na
predikciu cenovej kategórie
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 predpísaných liekov
- ▶ V tejto chvíli modely tréňované na malej anonymizovanej vzorke
- ▶ Kód pripravený na tréňovanie na veľkej vzorka na serveri

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

Tréňovanie,
testovanie a
validácia

Tréňovanie
Tréňovanie modelu na
predikciu budúcich
záznamov
Tréňovanie modelu na
predikciu cenovej kategórie
Validácia

Ďalšie kroky

Údaje pre časový rad

- ▶ Čas od narodenia pacienta
- ▶ Medicínsky výkon
- ▶ Liek
- ▶ Diagnóza
- ▶ 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
Trénovanie modelu na
predikciu budúcich
záznamov
Trénovanie modelu na
predikciu cenovej kategórie
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
Trénovanie modelu na predikciu budúcich záznamov
Trénovanie modelu na predikciu cenovej kategórie
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

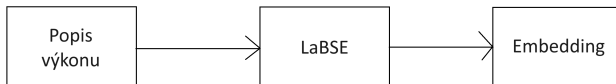
Trénovanie,
testovanie a
validácia

Trénovanie a testovanie
Trénovanie modelu na
predikciu budúcich
záznamov
Trénovanie modelu na
predikciu cenovej kategórie
Validácia

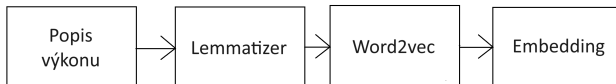
Ďalšie kroky

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

Trénovanie modelu na predikciu budúcich záznamov

Trénovanie modelu na predikciu cenovej kategórie

Validácia

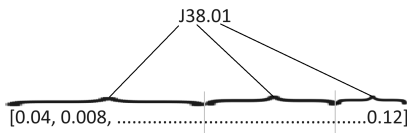
Ďalšie kroky

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

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

Trénovanie modelu na
predikciu budúcich
záznamov

Trénovanie modelu na
predikciu cenovej kategórie

Validácia

Ďalšie kroky

Lieky

► 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

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
Trénovanie modelu na
predikciu budúcich
záznamov

Trénovanie modelu na
predikciu cenovej kategórie
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

Trénovanie modelu na predikciu budúcich záznamov

Trénovanie modelu na predikciu cenovej kategórie

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

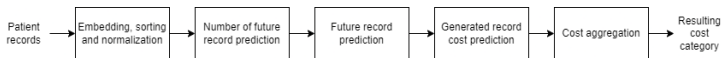
Trénovanie modelu na
predikciu budúcich
záznamov

Trénovanie modelu na
predikciu cenovej kategórie
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

Trénovanie modelu na predikciu budúcich záznamov

Trénovanie modelu na predikciu cenovej kategórie

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

Trénovanie modelu na
predikciu budúcich
záznamov

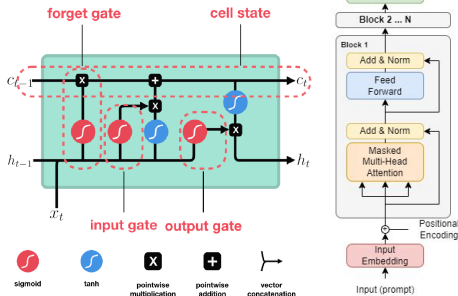
Trénovanie modelu na
predikciu cenovej kategórie

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

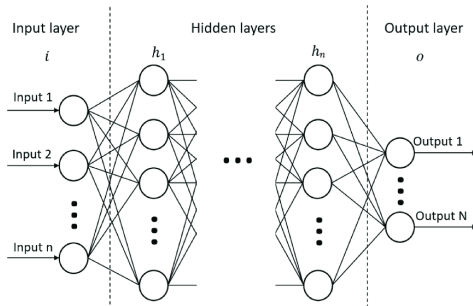
Trénovanie a testovanie
Trénovanie modelu na
predikciu budúcich
záznamov

Trénovanie modelu na
predikciu cenovej kategórie
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
Trénovanie modelu na
predikciu budúcich
záznamov
Trénovanie modelu na
predikciu cenovej kategórie
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

Trénovanie modelu na predikciu budúcich záznamov

Trénovanie modelu na predikciu cenovej kategórie

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

Trénovanie modelu na
predikciu budúcich
záznamov

Trénovanie modelu na
predikciu cenovej kategórie

Validácia

Ďalšie kroky

Trénovanie a testovanie

- ▶ Predikčné modely sa trénujú a testujú samostatne
- ▶ Chyba predikcie záznamu - váhovaná euklidovská vzdialenosť od reálneho záznamu
- ▶ Chyba predikcie cenovej kategórie - 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

Trénovanie modelu na
predikciu budúcich
záznamov
Trénovanie modelu na
predikciu cenovej kategórie
Validácia

Ďalšie kroky

Trénovanie modelu na predikciu budúcich záznamov

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

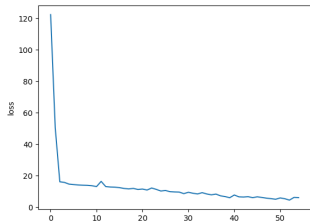
Trénovanie a testovanie

Trénovanie modelu na predikciu budúcich záznamov

Trénovanie modelu na predikciu cenovej kategórie
Validácia

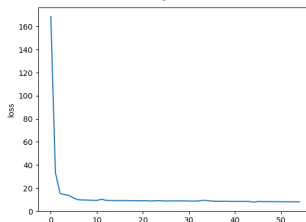
Ďalšie kroky

LSTM model



- ▶ Trénovacia chyba: 6.11
- ▶ Testovacia chyba: 10.21

Decoder-only transformer



- ▶ Trénovacia chyba: 8.17
- ▶ Testovacia chyba: 9.69

Trénovanie modelu na predikciu budúcich záznamov

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

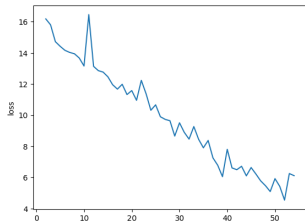
Trénovanie a testovanie

Trénovanie modelu na predikciu budúcich záznamov

Trénovanie modelu na predikciu cenovej kategórie
Validácia

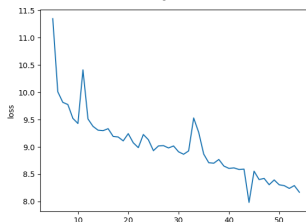
Ďalšie kroky

LSTM model



- ▶ Trénovacia chyba: 6.11
- ▶ Testovacia chyba: 10.21

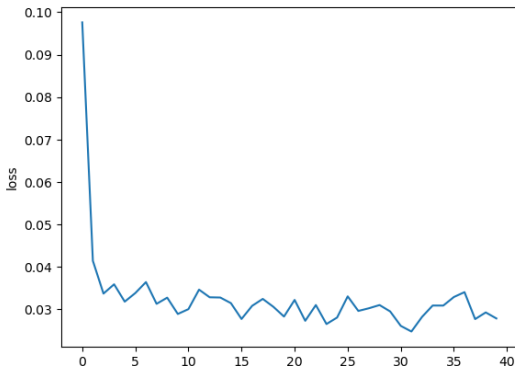
Decoder-only transformer



- ▶ Trénovacia chyba: 8.17
- ▶ Testovacia chyba: 9.69

Trénovanie modelu na predikciu cenovej kategórie

MLP model



- ▶ Trénovacia chyba: 0.028
- ▶ Testovacia chyba: 0.029
- ▶ Presnosť: 80.4%

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

Trénovanie modelu na
predikciu budúcich
záznamov

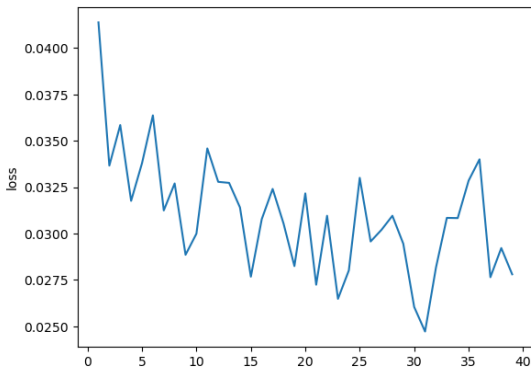
Trénovanie modelu na
predikciu cenovej kategórie

Validácia

Ďalšie kroky

Trénovanie modelu na predikciu cenovej kategórie

MLP model



- ▶ Trénovacia chyba: 0.028
- ▶ Testovacia chyba: 0.029
- ▶ Presnosť: 80.4%

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
Trénovanie modelu na
predikciu budúcich
záznamov
Trénovanie modelu na
predikciu cenovej kategórie
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
5. Predbežné výsledky:
 - 5.1 Správna kategória: 22%
 - 5.2 Blízka kategória: 34%
 - 5.3 Nesprávna kategória: 44%

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
Trénovanie modelu na
predikciu budúcich
záznamov
Trénovanie modelu na
predikciu cenovej kategórie
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

Trénovanie modelu na predikciu budúcich záznamov

Trénovanie modelu na predikciu cenovej kategórie

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

Trénovanie modelu na
predikciu budúcich
záznamov

Trénovanie modelu na
predikciu cenovej kategórie

Validácia

Ďalšie kroky

Ďalšie kroky

- ▶ Trénovanie, testovanie a nastavovanie hyperparametrov modelov na väčšom datasete na 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
Trénovanie modelu na
predikciu budúcich
záznamov
Trénovanie modelu na
predikciu cenovej kategórie
Validácia

Ďalšie kroky