Predikcia budúcich nákladov na pacienta

Marián Kravec školiteľ: MSc. František Dráček

Predikcia budúcich nákladov na pacienta

Cieľ práce

Data

Embedo

Diagnóz

Predikcia náklado

Proces prodikcie

Výpočet počtu budúcic

záznamov Predikcia budúcich

záznamov Predikcia cenovej kategój

Predikcia cenovej kategóri záznamu

Trénovanie, testovanie a

alidácia

Trénovanie a testovanie

Trénovanie modelu na predikciu budúcich

áznamov

Trénovanie modelu na predikciu cenovej kateg

unduoru

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

- mala a d

mbedaing /ýkonv

Diagnôzy Lieky

Predikcia nákladov

Proces predikcie Výpočet počtu budúc

raznamov Predikcia budúcich

iznamov redikcia cenovej kateg

namu

testovanie a

Trénovanie a testovanie

Trénovanie modelu na

predikciu budúcich záznamov

Trénovanie modelu i predikciu cenovej ka

Dalšie krokv

Daiolo Riorty

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



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

-...

Výkony

Diagnózy Lieky

Predikcia nákladov

Proces predikcie Výpočet počtu budúcich záznamov

zaznamov Predikcia budúcich záznamov

edikcia cenovej kateg znamu

Trénovanie, testovanie a

Trénovanie a testovanie

Trénovanie modelu na predikciu budúcich

Trénovanie modelu predikciu cenovej k

S-1X!- I---I--



Dáta

- Zdroj dát: NCZI
- Typ dát: poistné dáta
- Jeden záznam: informácia o jednom výkone/lieku konkrétneho pacienta
- Datasety:
 - Dataset ambulantnej zdravotnej starostlivosti
 - Dataset predpísaných liekov
- V tejto chvíli modely trénované na malej anonymizovanej vzorke
- Kód pripravený na trénovanie na veľkej vzorka na serveri

Predikcia budúcich nákladov na pacienta

Dáta



Údaje pre časový rad

- Čas od narodenia pacienta
- Medicínsky výkon
- Liek
- Diagnóza
- Cena výkonu/lieku

Predikcia budúcich nákladov na pacienta

Dáta



Embedding

Predikcia budúcich nákladov na pacienta

Embeddina



Výkony

Language-agnostic BERT sentence embedding model (LaBSE)



Lemmatizer + Word2vec model



Predikcia budúcich nákladov na pacienta

Výkony

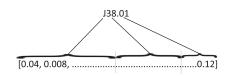


Diagnózy

► ICD-10-CM kódy

[znak][dvojčíslie].[dvojčíslie]

- 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

ieľ práce

ata

mbedding /ýkony

Diagnózy

Predikcia nákladov

Proces predikcie Výpočet počtu bud

záznamov Predikcia hudúcich

záznamov Predikcia cenovej kategó

énovanie,

estovanie a validácia

Trénovanie a testovanie

Trénovanie modelu predikciu budúcich záznamov

Trénovanie modelu predikciu cenovej ka

Validácia



Lieky

ATC kód

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

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

Predikcia budúcich nákladov na pacienta

Liekv



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

Embodo

ýkony

Diagnózy Lieky

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich

záznamov Predikcia budúcich

edikcia cenovej kateg

énovanie,

testovanie a validácia

Trénovanie a testovanie

Trénovanie modelu na predikciu budúcich

záznamov Trénovanie modelu predikciu cenoval l

predikciu cenovej ka Validácia



Proces predikcie

- 1. Načítanie dát pacienta, embedding a normalizácia
- Výpočet počtu záznamov na budúci rok
- Predikcia budúcich záznamov
- 4. Predikcia cenových kategórii budúcich záznamov
- 5. Výpočet celkovej cenovej kategórie



Predikcia budúcich nákladov na pacienta

Proces predikcie



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

Data

Embedd

Diagnózy

Predikcia nákladov

Proces predikcie

Výpočet počtu budúcich

záznamov

záznamov

Predikcia cenovej kategóri

Trénovanie, testovanie a

Validacia

Trénovanie a testovanie

predikciu budúcich záznamov

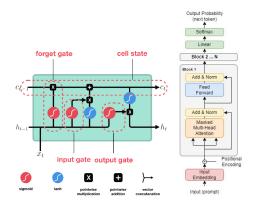
Trénovanie modelu na predikciu cenovej kate

-1¥!--1...-1...



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

Jata

Embedo

Výkony Diagnózy

Jiagnozy Lieky

redikcia nákladov

Proces predikcie Výpočet počtu bi

Výpočet počtu budí záznamov Predikcia budúcich

záznamov Predikcia cenovei kate

áznamu

estovanie a

Trénovanie a testovar

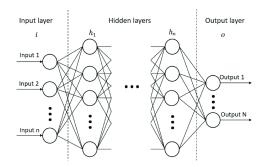
záznamov Trénovanie modeli

rénovanie modelu n vredikciu cenovej kal



Predikcia cenovej kategórie záznamu

- Namiesto konkrétnej ceny predikujeme cenovú kategóriu
- Použitý model: Multilayer perceptron



Predikcia budúcich nákladov na pacienta

Predikcia cenovei kategórie záznamu



Cieľ práce

Dáta

Embedding

Výkony Diagnóz Liekv

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

ieľ práce

Dáta

Embedo

Výkony Diagnózy

Prodikcia nákladov

Proces predikcie Výpočet počtu budúc

Predikcia budúcich záznamov

edikcia cenovej kateg znamu

Trénovanie, testovanie a validácia

Trénovanie a testovanie

Trénovanie modelu r predikciu budúcich

Trénovanie modelu predikciu cenovej k

) alšie kroky

Daisie 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

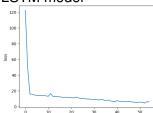
Predikcia budúcich nákladov na pacienta

Trénovanie a testovanie



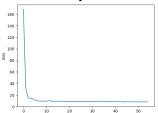
Trénovanie modelu na predikciu budúcich záznamov

LSTM model



- Trénovacia chyba: 6.11
- ► Testovacia chyba: 10.21

Decoder-only transformer



- ▶ Trénovacia chyba: 8.17
- Testovacia chyba: 9.69

Predikcia budúcich nákladov na pacienta

nei prace

Jata

mpea

Diagnózy Lieky

Predikcia nákladov

Vinoset post

záznamov

redikcia budúcich áznamov

edikcia cenovej kategór znamu

testovanie a

alidacia

Irenovanie a testovanie

Trénovanie modelu na predikciu budúcich

Trénovanie modelu na predikciu cenovej kategórie

Ďalčio kroky

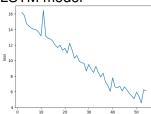
Ďalšie kroky

záznamov



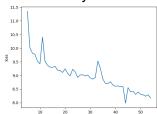
Trénovanie modelu na predikciu budúcich záznamov

LSTM model



- Trénovacia chyba: 6.11
- Testovacia chyba: 10.21

Decoder-only transformer



- ▶ Trénovacia chyba: 8.17
- Testovacia chyba: 9.69

Predikcia budúcich nákladov na pacienta

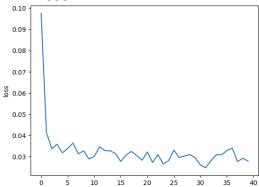
Trénovanie modelu na predikciu budúcich

záznamov



Trénovanie modelu na predikciu cenovej kategórie

MLP model



► Trénovacia chyba: 0.028

Testovacia chyba: 0.029

Presnosť: 80.4%

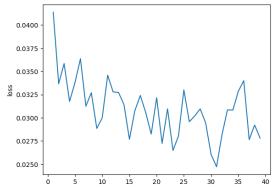
Predikcia budúcich nákladov na pacienta

Trénovanie modelu na predikciu cenovei kategórie



Trénovanie modelu na predikciu cenovej kategórie

MLP model



▶ Trénovacia chyba: 0.028

Testovacia chyba: 0.029

Presnosť: 80.4%

Predikcia budúcich nákladov na pacienta

Diel práce

Dáta

Embed

Výkony

Diagnózy Liekv

Predikcia nákladov

Proces predikcie

Výpočet počtu budú záznamov

Predikcia budúcich záznamov

redikcia cenovej kategó aznamu

testovanie a

Trénovanie a 1

Trénovanie modelu

Trénovanie modelu na predikciu cenovej kategórie

Validácia



Validácia

- Rozdelenie záznamov pacienta na posledný rok a staršie
- Výpočet skutočnej celkovej cenovej kategórie pacienta v poslednom roku
- Predikcia počtu záznamov, jednotlivých záznamov, ich cenovej kategórie a celkovej cenovej kategórie
- 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

Jieľ práce

Dáta

Embedd

Výkony

Diagnózy Lieky

Predikcia nákladov

Proces predik Výpočet počtu

záznamov

iznamov

redikcia cenovej kateg áznamu

Trénovanie, testovanie a

/alidacia

Trénovanie a t

edikciu budúcich

Trénovanie mod

edikciu cenovej ka

Validácia



Výpočet počtu budúcich záznamov

Ďalšie kroky

Predikcia budúcich nákladov na pacienta



Ďalšie kroky

- Trénovanie, testovanie a nastavovanie hyperparametrov modelov na väčšom datasete na serveri
- Výber najlepších hyperparametrov
- Spísanie výsledkov

Predikcia budúcich nákladov na pacienta

