

**Inovácia softvérového produktu pre oblasť zdravotného poistenia využitím
metód strojového učenia
Akronym: SU4ZP**

VÝROČNÁ SPRÁVA ZA ROK 2020

Kód projektu: 313012S703

Prijímateľ: VSL Software a.s.

Partner1: Technická univerzita v Košiciach

Partner 2: Univerzita Pavla Jozefa Šafárika v Košiciach

Dátum začatia a ukončenia projektu: 01/2020 - 6/2022

Zazmluvnená výška NFP pre UPJŠ: 71 299,53 Eur

Žiadaná výška NFP za rok 2020: 4 967,82 Eur

Preplatená výška NFP v roku 2020: 11 297,37 Eur

Názov hlavnej aktivity, na realizácii ktorej sa UPJŠ v Košiciach podieľa: Priemyselný výskum

Garantom aktivity partnera UPJŠ v Košiciach je RNDr. Erik Bruoth, PhD. z centra informatiky a informačných technológií TIP UPJŠ

Konkrétna činnosť/podaktivita, na realizácii ktorej sa partner podieľa:

Míľnik M1:

Formulácia úloh strojového učenia potrebných pre dosiahnutie požadovaných biznis cieľov:

V rámci úlohy sme sa detailne oboznamovali s doménou zdravotného poistenia a s dostupnou dátovou. Na základe formulovaných otázok z biznis prostredia boli identifikované oblasti, v ktorých je vhodné aplikovať metódy strojového učenia vzhľadom na dostupnosť a kvalitu údajov. Primárne sme sa zameriavalí na oblasť vykazanej zdravotnej starostlivosti s cieľom identifikovať anomálie resp. podobnosti v zaužívaných postupoch vykazovania. Ďalšia oblasť, ktorá bola identifikovaná, sa týka možnosti predikcie očakávanej zdravotnej starostlivosti vzhľadom na vykázané záznamy. Otvorila sa otázka možnosti automatizácie spracovania zdravotných záznamov pri skriningových vyšetreniach so širokým záberom poistencov. Zapojením strojového učenia v tejto oblasti má potenciál zásadným spôsobom zefektívniť skriningové vyšetrenia a zvýšiť dopad na zdravie spoločnosti. Zameriavalí sme sa na jasnú a precíznu formuláciu úloh pre strojové učenie aj so špecifikovaním metodiky vyhodnocovania. Zásadným spôsobom do týchto procesov ale vstupovala kvalita dát a ich systematické odchýlky spôsobené implementovanými procesmi jednotlivých poskytovateľov zdravotnej starostlivosti resp. legislatívou. Boli navrhnuté a implementované štatistické metódy, tak aby sme niektoré z týchto odchýliek dokázali kompenzovať na únosnú mieru. Primárne sa jednalo o bodové a skupinové anomálie.

Takto sa podarilo pripraviť revidovanú dátovú sadu pre potreby strojového učenia v niektorých spomínaných oblastiach.

Kritický prieskum existujúcich metód:

Vzhľadom na to, že nie sme priekopníkmi v implementácii metód strojového učenia v tejto doméne, zamerali sme sa v tejto úlohe na analýzu publikovaných prípadových štúdií. Vzhľadom na množstvo a rôznu kvalitu aj v recenzovaných časopisoch, nebolo úplne triviálne identifikovať štúdie, ktoré by bolo užitočné overiť nad našou dátovou sadou. Mnoho krát metódy neboli použité úplne korektne resp. nezohľadňovali charakter našej dátovej sady. Niektoré implementácie neboli celkom jasne popísané a aj napriek snahe o vyjasnenie kritických bodov s autormi sme nedospeli ku presvedčeniu o ich vhodnosti.

Výber vhodných metód a ich úpravy:

Identifikovali sme teda aj metódy z iných doménových oblastí, ktoré by mohli byť použiteľné v nami formulovaných úlohách. Hlavne sa jednalo o grafové algoritmy z oblasti analýzy veľkých sietí (založené na náhodných prechádzkach resp. spektrálnych analýzach), resp. z oblasti spracovania prirodzeného jazyka (napr. pre učenie bez učiteľa). Niektoré doteraz používané algoritmy sme boli nútení modifikovať, ako napr. SPADE pre vyhľadávanie frekventovaných položiek tak, aby boli použiteľné v našich úlohách.

Vytvorenie modelov strojového učenia a vyhodnotenie:

Z pozície partnera sme z identifikovaných metód vytvárali prototypy pre overenie ich použitia nad reálnymi dátami a vyhodnotenie. Mnohé prototypy, ale nepotvrdili svoj potenciál alebo skôr poukázali na komplikovanosť implementácie vzhľadom na relatívny prínos k riešeniu úloh. Niektoré prístupy budú ďalej skúmané v druhom roku riešenia. Prístupy na báze náhodných lesov, napr. pre identifikáciu podozrivých výkazov zdravotnej starostlivosti, boli implementované partnerom v rámci produktu. Niektoré sú vo fáze optimalizácie algoritmov, napr. GloVe reprezentácie. V prípade ak sa podarí navrhnúť vhodnú úpravu, pristúpi sa ich implementácii v priebehu ďalšieho obdobia.