



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

### **VÝROČNÁ SPRÁVA ZA ROK 2021**

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 2021: 25 129,93 Eur ( z toho PF: 11 036,47 Eur)

Preplatená výška NFP v roku 2021: 16 278,41 Eur ( z toho PF: 6 758,33 Eur)

Názov hlavnej aktivity, na realizácii ktorej sa UPJŠ v Košiciach podieľa: Priemyselný výskum  
Garantom aktivity partnera UPJŠ v Košiciach bol do 30.11.2021 RNDr. Erik Bruoth, PhD. z centra informatiky a informačných technológií TIP UPJŠ, od 1.12.2021 je to prof. RNDr. Ivan Žežula, CSc. z ústavu matematických vied PF UPJŠ.

Konkrétna činnosť/podaktivita, na realizácii ktorej sa partner podieľal v roku 2021:

Míľnik M1:

1.1 Formulácia úloh strojového učenia potrebných pre dosiahnutie požadovaných biznis cieľov:

V rámci tejto podaktivity sme naše úlohy v roku 2021 sformulovali vo forme skúmania metód a algoritmov strojového učenia a ich použitia v oblastiach zdravotného poistenia a zdravotnej starostlivosti. Cieľom bolo tieto metódy a algoritmy analyzovať na základe formulovaných cieľov z biznis prostredia a tiež skúmať možnosti ich aplikácie na dostupných údajových množinách. Významnou úlohou bolo dostupné údajové množiny vhodne pripraviť, predspracovať a analyzovať pre ich ďalšie použitie. Primárne sme sa zameriavali na oblasť EKG záznamov s cieľom identifikovať vybrané typy ochorení srdca na týchto záznamoch. Okrem tohto hlavného cieľa sme sformulovali aj úlohu optimálnej segmentácie EKG záznamov, ktorej riešenie môže pomôcť ako vstup pre ďalšie analýzy v tejto oblasti. Významnou úlohou v roku 2021 bolo dané metódy a algoritmy navrhnuť a implementovať.

1.4 Kritický prieskum existujúcich metód:

Pri analýze publikovaných prípadových štúdií v zahraničnej literatúre sme identifikovali hlavné výhody a nevýhody jednotlivých prístupov v oblasti strojového učenia a možnosti ich použitia pri riešení našich úloh v oblasti EKG záznamov. Zistili sme, že viaceré metódy boli overované na špeciálnych údajových množinách a nemusia byť univerzálne použiteľné. Po získaní prehľadu literatúry sme navrhli rôzne modifikácie a úpravy existujúcich metód za účelom dosiahnutia stanovených cieľov.



### 1.5 Výber vhodných metód a ich úpravy:

Identifikovali sme ďalšie metódy z iných doménových oblastí, ktoré vykazovali potenciál pre použitie v nami formulovaných úlohách. V roku 2021 sme sa zaoberali najmä metódami nekontrolovaného učenia alebo polokontrolovaného učenia. Pri týchto typoch úloh strojového učenia sú veľmi dôležité augmentačné metódy, pomocou ktorých môžeme analyzovať aj údajové sady bez označení, teda bez hodnôt cieľového atribútu. Na týchto metódach je založené aj kontrastívne učenie z oblasti analýzy obrazu, ktoré sme tiež aplikovali. Viaceré existujúce metódy a algoritmy sme modifikovali pre použiteľnosť v našich úlohách. Pre úlohy segmentácie EKG záznamov sme vybrali aj viaceré metódy z oblasti spracovania prirodzeného jazyka a analýzy obrazu.

### 1.6 Vytvorenie modelov strojového učenia a vyhodnotenie:

Z pozície partnera projektu sme z identifikovaných metód vytvárali prototypy pre overenie ich použitia nad reálnymi údajmi. Súčasťou tejto časti bolo aj vyhodnotenie a porovnanie výsledkov rôznych metód, najmä v oblasti neurónových sietí, rozhodovacích stromov alebo ďalších štatistických metód. Dôležité postavenie v našich modeloch mali konvolučné neurónové siete so zapracovaním ďalších prvkov a mechanizmov. Skúmané boli aj prístupy založené na metódach náhodných lesov. Niektoré ďalšie prístupy a metódy budú ďalej skúmané v treťom roku riešenia. Pri segmentácii EKG záznamov sme použili aj špeciálnu stratovú funkciu z oblasti spracovania prirodzeného jazyka.