

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Odhad kompatibility dárce a příjemce pro transplantaci ledvin pomocí strojového učení
Jméno autora:	Matěj Klouček
Typ práce:	bakalářská práce
Fakulta:	Fakulta jaderná a fyzikálně inženýrská (FJFI)
Katedra:	Katedra matematiky
Oponent práce:	doc. RNDr. Ľubomír Antoni, PhD.
Pracoviště oponenta práce:	Univerzita Pavla Jozefa Šafárika v Košiciach, Prírodovedecká fakulta, Ústav informatiky

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
<p>Bakalářska práca sa venuje aktuálnej a prínosnej problematike strojového učenia v oblasti zdravotníctva, ktorej sa v súčasnosti intenzívne venujú aj na popredných svetových univerzitách a v ďalších významných IT spoločnostiach. Cieľom práce bolo navrhnúť algoritmus predikujúci kompatibilitu darcu a príjemcu obličiek a overiť jeho presnosť na reálnych a vygenerovaných dátach. Zároveň bolo cieľom porovnať presnosť navrhnutého algoritmu na dátach z rôznych vzoriek populácie a prípadne navrhnúť jeho úpravy. Sformulované zadanie bakalárskej práce rieši veľmi aktuálne a prínosné výzvy súčasnej vedy.</p>	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
<p>Stanovené ciele práce sa podarilo naplniť výborne. Práca obsahuje výsledky vlastnej odbornej a tvorivej činnosti autora vrátane hodnotných implementačných výsledkov. Hlavným prínosom práce je zostavenie modelu náhodných lesov s využitím analýzy prežitia, pomocou ktorého autor dosiahol vyšší C-index ako pri použití Coxovho regresného modelu na predikciu kompatibility darcu a príjemcu obličiek. Autor navyše zistil vplyv veku príjemcu, či HLA nezhody na prežívanie štepu. V závere praktickej časti autor skúma aj vplyv etnicity (10 etnických skupín) na výsledok transplantácie.</p>	

Zvolený postup řešení	vhodný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
<p>Autor vo svojom riešení použil vhodné štatistické metódy a metódy strojového učenia, pričom v dostatočnej miere zdôvodnil ich použitie. Pri riešení stanovených cieľov najprv pracoval s dátami z Institutu klinickej a experimentálnej medicíny v Prahe. Na zostavenie modelov strojového učenia používal aj rozsiahle dáta z neziskovej organizácie UNOS, ktorá riadi systém transplantácie orgánov v Spojených štátoch amerických z obdobia rokov 1984 až 2022. Na analýzu dát použil rozšírenie metódy náhodných lesov pomocou analýzy prežitia. Pomocou tejto metódy zostavil model na ohodnotenie kompatibility medzi darcom a príjemcom predikovaním času prežitia založeného na predtransplantačných metrikách. Vyhodnotenie modelu vykonal pomocou Harrellovho C-indexu. Model následne použil aj na skúmanie rôznych faktorov ovplyvňujúcich prežívanie štepu. Jednotlivé fázy analýzy údajov sú spracované odborným a systematickým spôsobom.</p>	

Odborná úroveň

výborná

Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.

Použité výrazové prostriedky a pojmy sú na kvalitnej odbornej úrovni, zodpovedajú terminológii používanej v odbornej literatúre. Teoretická časť práce je súčasťou prvých štyroch kapitol. Autor v nich popisuje metódy strojového učenia, analýzu prežitia, náhodné lesy založené na analýze prežitia a transplantáciu obličiek. Autor v prvej kapitole popisuje typy modelov strojového učenia a postup pri budovaní týchto modelov. V podkapitole 1.1.1 by bolo vhodné spomenúť aj metódy tzv. samokontrolovaného učenia (self-supervised learning), ktoré sa v posledných rokoch dostávajú do popredia najmä pri analýze veľkých neoznačkovaných dátových sád. V podkapitole 1.2.1 pri rozdelení dát je vhodné spomenúť aj krížovú validáciu (cross-validation), ktorá umožňuje budovať model strojového učenia iteratívne na rôznych podmnožinách dát. Keďže veľkou výhodou rozhodovacích stromov je vizualizácia vybudovaného modelu, ocenil by som v teoretickej alebo praktickej časti obrázok, ktorý by znázorňoval takýto strom na konkrétnom príklade.

Formální a jazyková úroveň

výborná

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.

Bakalárska práca je napísaná v anglickom jazyku na kvalitnej jazykovej úrovni. Práca v rozsahu 54 strán je rozdelená do siedmich hlavných kapitol a je doplnená úvodom a záverom. Jednotlivé kapitoly a podkapitoly sú vnútorne previazané a tvoria zmysluplný celok. Súčasťou práce je viacero tabuliek a grafov, ktoré vhodne dopĺňajú predkladaný text. V práci sa nachádzajú len malé formálne nedostatky, napríklad:

- strana 12: v zápise (x_1, \dots, x_n) chýba čiarka pred x_n ,
- strana 12: symbol \hat{n} nie je v texte definovaný (množina všetkých indexov)
- strana 30: skratky TP, TPR, FP, ... by mali ostať aj vo vzorcoch bez kurzívy, keďže takto boli definované v texte

Výběr zdrojů, korektnost citací

výborné

Vyjáďřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Zoznam použitej literatúry je relevantný, citovanie uvedenej literatúry je správne a v celej práci konzistentné. Autor pracoval aj s aktuálnymi literárnymi zdrojmi z posledného obdobia. Formát jednotlivých bibliografických záznamov je v zozname použitej literatúry jednotný. Pomohlo by, ak by zoznam použitej literatúry bol buď v abecednom poradí alebo v poradí, v akom boli zdroje citované v texte.

Další komentáře a hodnocení

Vyjáďřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Zdrojový kód riešenia je dostupný vo forme Jupyter Notebooku, ktorý je vhodne štruktúrovaný pomocou nadpisov a podnadpisov a zodpovedá štandardným a korektným postupom v oblasti spracovania dát. Dosiahnuté závery môžu poslúžiť aj ako kvalitný základ pre ďalší vedecký výskum a experimenty.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Študent preukázal schopnosť integrovať teoretické a praktické vedomosti v skúmanej oblasti a tvorivo pristúpiť k riešeniu vybranej problematiky v oblasti strojového učenia a transplantácie obličiek. Počas obhajoby sa študent môže vyjadriť k nasledujúcim otázkam:

V kapitole 1.3 uvádzate, že rozhodovacie stromy sú väčšinou binárne a popisujete algoritmus ID3 pre prípad numerických vstupných atribútov (numerical features). Ako by sa zmenila situácia, ak by sme uvažovali v tomto algoritme kategoriálne vstupné atribúty (pohlavie, typ ochorenia) bez ich one-hot kódovania?

Vo svojej práci ste na optimalizáciu hyperparametrov použili metódu grid search. Akým spôsobom by bolo možné dosiahnuť efektívnejšiu časovú náročnosť Vášho riešenia?

Ako celok hodnotím bakalársku prácu Matěja Kloučka veľmi pozitívne a odporúčam ju prijať k obhajobe.

Predloženou závěrečnou prací hodnotím klasifikačním stupněm **A - výborně**.

Datum: 18.8.2023

Podpis:

