

ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

-обавезна садржина- свака рубрика мора бити попуњена

(сви подаци уписују се у одговарајућу рубрику, а назив и место рубрике не могу се мењати или изоставити)

I ПОДАЦИ О КОМИСИЈИ
<p>1. Датум и орган који је именовао комисију Декан Факултета техничких наука у Новом Саду на основу предлога матичне катедре и одлуке Наставно-научног већа Факултета техничких наука; Решење број: 012-199/50-2019</p> <p>2. Састав комисије са знаком имена и презимена сваког члана, звања, назива уже научне области за коју је изабран у звање, датума избора у звање и назив факултета, установе у којој је члан комисије запослен:</p> <p>ПРЕДСЕДНИК: Др Драган Иветић, редовни професор Примењене рачунарске науке и информатика; 13.1.2010.; Факултет техничких наука у Новом Саду</p> <p>ЧЛАН: Др Јелица Протић, редовни професор Рачунарска техника и информатика; 1.11.2017.; Електротехнички факултет у Београду</p> <p>ЧЛАН: Др Едита Стокић, редовни професор Ендокринологија, дијабетес и болести метаболизма; 1.3.2009.; Медицински факултет у Новом Саду</p> <p>ЧЛАН: Др Горан Сладић, ванредни професор Примењене рачунарске науке и информатика; 13.09.2016.; Факултет техничких наука у Новом Саду</p> <p>МЕНТОР: Др Александар Купусинац, ванредни професор Примењене рачунарске науке и информатика; 19.05.2016.; Факултет техничких наука у Новом Саду</p> <p>МЕНТОР: Др Ксенија Дорословачки, ванредни професор Теоријска и примењена математика; 8.7.2019.; Факултет техничких наука у Новом Саду</p>
II ПОДАЦИ О КАНДИДАТУ
<p>1. Име, име једног родитеља, презиме: Дуња (Вера) Врбашки</p> <p>2. Датум рођења, општина, држава: 09.10.1978. Нови Сад, Србија</p> <p>3. Назив факултета, назив студијског програма дипломских академских студија – мастер и стечени стручни назив Природно-математички факултет Нови Сад; Информатика – рачунарске науке, Мастер информатичар</p> <p>4. Година уписа на докторске студије и назив студијског програма докторских студија 2015; Рачунарство и аутоматика</p> <p>5. Назив факултета, назив магистарске тезе, научна област и датум одбране: /</p> <p>6. Научна област из које је стечено академско звање магистра наука: /</p>

III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

Примена машинског учења у проблему недостајућих података при развоју предиктивних модела

IV ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

Навести кратак садржај са назнаком броја страна, поглавља, слика, шема, графикона и сл.

Докторска дисертација кандидата Дуње Врбашки, под насловом **Примена машинског учења у проблему недостајућих података при развоју предиктивних модела**, је изложена на 112 страница. Текст дисертације је организован у 7 поглавља, списак литературе и 3 прилога у којима се, укупно, налази 25 слика, 14 табела и 3 листинга програмског кода. Списак коришћене литературе садржи 70 навода.

Садржај докторске дисертације је следећи:

1. Увод
2. Проблем недостајућих података и две научне заједнице
3. Подаци који недостају
4. Примена машинског учења у проблему недостајућих података
5. Подаци који недостају и развој предиктивних модела машинског учења
6. Студија случаја
7. Закључак

ПРИЛОГ

A1 Студија случаја: Експериментални резултати припреме АНН модела за импутацију

A2 Студија случаја: Експериментални резултати импутација

A3 Сагласност етичког одбора Клиничког центра Војводине за спровођење истраживања

V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

Поглавље 1: Увод

У уводном поглављу кандидат описује проблем недостајућих података и најчешће приступе његовом решавању ослањајући се на публиковане резултате светских истраживања. Указује на проблеме који се јављају уколико се приступи искључивању некомплетних података из процеса истраживања. Посебно се осврће на проблем који се јавља у домену примене машинског учења што представља идејни оквир истраживања докторске дисертације.

Након уводних разматрања кандидат јасно наводи предмет, хипотезе, циљеве и допринос истраживања.

Поглавље 2: Проблем недостајућих података и две научне заједнице

У овом поглављу кандидат разматра однос према проблему недостајућих података у оквирима две струке, односно научне заједнице које чине истраживачи из домена статистике и машинског учења што наводи и као инспирацију за рад на оваквом истраживању.

Поглавље 3: Подаци који недостају

У овом поглављу кандидат детаљно презентује проблем недостајућих података како га сагледавају статистичка истраживања инсистирајући на појмовима који су од интереса у дисертацији. Анализирани су процеси и методе импутације недостајућих података са посебним нагласком на оне који су најчешће коришћени у пракси укључујући и проблеме који се појављују његовом применом при развоју предиктивних модела машинског учења.

Поглавље 4: Примена машинског учења у проблему недостајућих података

У овом поглављу кандидат прелази са статистичких метода на методе машинског учења као потенцијалних метода импутације. Након прегледа основних појмова машинског учења и метода које су се користиле у изради експерименталног дела истраживања, кандидат даје преглед литературе која се, овај пут, односи на примену машинског учења код проблема недостајућих података посебно се осврћући на истраживања која се баве импутацијом што је у складу са циљевима рада.

У посебном делу поглавља, кандидат даје преглед остварених истраживачких резултата који су прихваћени за публикацију у међународном часопису. Студија се односи на прву хипотезу истраживања докторске дисертације и показује да се вештачке неуралне мреже могу искористити за импутацију лабораторијских вредности при изградњи предиктивних модела кардио-метаболичког ризика заснованих на једноставним параметрима.

Поглавље 5: Подаци који недостају и развој предиктивних модела машинског учења

За разлику од претходних поглавља, овде кандидат разматра другу страну предмета истраживања: како недостајући подаци утичу на сам процес изградње предиктивног модела у домену примене машинског учења. На почетку поглавља се управо истиче значај разматрања проблема недостајућих података приликом развоја модела. Након тога, кандидат, прво, наводи опште поступке и термине као што су унакрсне валидације и мере које се често користе приликом изградње предиктивних модела машинског учења. Посебно се осврће на процес аутоматског генерисања синтетичких недостајућих података који је неопходан корак у експерименталним симулацијама. Након тога, дефинише предлог две методологије које будући истраживачи могу користити приликом анализе утицаја недостајућих података на перформансе својих модела. Формулише два таква алгорита у складу са различитом наменом и циљевима коришћења: када је потребно извршити импутацију постојећих, конкретних недостајућих података и када се врши општа анализа потенцијалног утицаја постојања недостајућих података приликом развоја модела. Уз сваки алгоритам су дата јасна упутства и значајне напомене које могу усмерити дизајн експеримента и ток истраживања.

Поглавље 6: Студија случаја

Ово поглавље представља експерименталне резултате истраживања пратећи претходно дефинисану методологију. Испитује се утицај импутација преко вештачких неуралних мрежа недостајућих података при изградњи предиктивног модела за процену метаболичког синдрома реализованог алгоритмом случајних шума (енг. random forest) публикованог у реномираном међународном часопису. Кандидат прво представља објављене резултате који се односе на предиктивни модел од интереса. Након тога се описује проблем и дају се информације о коришћеним подацима, методама и поставкама као и напомене о посебној нотацији која се користи за представљање посматраних случајева. Посебно је издвојена припрема за анализу где се врше експерименти за избор хиперпараметара метода које ће се касније користити у самим процесима импутације.

Извршено је испитавање већег броја сценарија недостајућих података где је посебно раздвојена анализа импутација лабораторијских вредности у односу на импутацију вредности код предиктора. За сваки сценарио недостајања, резултати су приказани табеларно и графички и извршена је дискусија резултата за сваки од њих посебно. На крају, кандидат даје општу дискусију резултата и ограничења и утврђује да вештачке неуралне мреже заиста успешно врше импутацију података што резултира бољим перформансама финалног модела у односу на моделе развијане над комплетним скупом.

Поглавље 7: Закључак

У последњем поглављу су сумирани резултати и изнета закључна разматрања дисертације уз потврђивање оствареног доприноса.

Наведени су различити правци наредних истраживања и изнета је теза да се, уз помоћ модерних сазнања, можда може формирати нови оквир за посматрање проблема недостајућих података у односу на устаљене принципе.

Литература

Дата је листа коришћене литературе која садржи 70 навода.

Додатак

У прилогу су приказане слике и табеле које се односе на експерименталне резултате као и документ о сагласности етичког одбора за коришћење података са клинике за ендокринологију, дијабетес и болести метаболизма Клиничког центра Војводине.

VI СПИСАК НАУЧНИХ И СТРУЧНИХ РАДОВА КОЈИ СУ ОБЈАВЉЕНИ ИЛИ ПРИХВАЋЕНИ ЗА ОБЈАВЉИВАЊЕ НА ОСНОВУ РЕЗУЛТАТА ИСТРАЖИВАЊА У ОКВИРУ РАДА НА ДОКТОРСКОЈ ДИСЕРТАЦИЈИ

Таксативно навести називе радова, где и када су објављени. Прво навести најмање један рад објављен или прихваћен за објављивање у часопису са ISI листе односно са листе министарства надлежног за науку када су у питању друштвено-хуманистичке науке или радове који могу заменити овај услов до 01.јануара 2012. године. У случају радова прихваћених за објављивање, таксативно навести називе радова, где и када ће бити објављени и приложити потврду о томе.

Vrbaški D., Vrbaški M., Kupusinac A., Ivanović D., Stokić E., Ivetić D., Doroslovački K.: Methods for Algorithmic Diagnosis of Metabolic Syndrome, Artificial Intelligence in Medicine, 2019, Vol. 101, ISSN 0933-3657, DOI: <https://doi.org/10.1016/j.artmed.2019.101708> [M21]

Vrbaški D., Kupusinac A., Doroslovački R., Stokić E., Ivetić D.: Missing data imputation in cardiometabolic risk assessment: A solution based on artificial neural networks, Computer Science and Information Systems, 2020. forthcoming [M23]

Kupusinac A., Stokić E., Doroslovački R., Ivetić D., **Vrbaški D.**, Vrbaški M., Ivanović D.: The Detection of Low HDL Cholesterol Level by Using Artificial Neural Networks, PSU-UNS International Conference on Engineering and Technology – ICET, 2017 [M33]

Kupusinac A., Stokić E., Okanović A., Vasiljević M., Stojić I., **Vrbaški D.**: Prediction of the HDL Cholesterol Level by Using Artificial Neural Networks, International Conference Industrial Engineering and Environmental Protection – IZS, 2015 [M33]

VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА

Истраживање приказано у дисертацији је довело до следећих конкретних закључака који се односе на постављене хипотезе истраживања:

- Методе машинског учења се могу разматрати као методе импутације недостајућих података. Конкретно, показано је да вештачка неурална мрежа остварује боље резултате као метод импутације него друге једноструке методе код лабораторијских података који се користе за процену кардио-метаболичког ризика. Ови резултати су прихваћени за објављивање у међународном часопису.
- Проблему недостајућих података се мора приступати са пажњом приликом развоја предиктивних модела машинског учења. Могуће је дефинисати методологију која

обезбеђује истраживачима, ван домена статистике где се изучава теорија недостајућих података, да формулишу и прецизно реализују експерименте који ће им омогућити анализу утицаја недостајућих података при изградњи предиктивних модела машинског учења.

- Применом наведене методологије је утврђено да се вештачке неуралне мреже могу искористити као метод импутације недостајућих података приликом развоја модела предикције постојања метаболичког синдрома и то недостајућих података и код предиктора али и код вредности које учествују само у фази претпроцесирања.

VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА

Експлицитно навести позитивну или негативну оцену начина приказа и тумачења резултата истраживања.

Дисертација јасно наводи предмет, хипотезе, циљеве и доприносе истраживања. Такође, изнете методе, резултати и њихова дискусија су наведени на јасан и концизан начин. У оквиру истраживања су коришћене адекватне и потврђене начне методе. Кандидат је успешно потврдио све постављене хипотезе, одговорио на постављена истраживачка питања и остварио циљеве истраживања.

На основу ових показатеља комисија даје **позитивну оцену начина приказа и тумачења резултата истраживања.**

Рад је проверен у претраживачу Google, тако што су сви делови текста дисертације проверавани да ли постоји плагијаризам и утврђено је да докторска дисертација кандидата ХУ представља оригинални рукопис. **НАПОМЕНА:** Овакав начин провере на плагијаризам је урађен, имајући у виду чињеницу да софтвер за детекцију плагијаризма iThenticate није био на располагању због недостатка лиценци за његово коришћење.

IX КОНАЧНА ОЦЕНА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

Експлицитно навести да ли дисертација јесте или није написана у складу са наведеним образложењем, као и да ли она садржи или не садржи све битне елементе. Дати јасне, прецизне и концизне одговоре на 3. и 4. питање:

1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме

Докторска дисертација је **написана у складу са образложењем** које је наведено у пријави теме.

2. Да ли дисертација садржи све битне елементе

Докторска дисертација својим насловом, садржајем, избором тема, оригиналним резултатима истраживања, начином тумачења и применом тих резултата **садржи све битне елементе** који се захтевају за радове овакве врсте.

3. По чему је дисертација оригиналан допринос науци

Разматрајући целокупну материју поднете докторске дисертације кандидата Дуње Врбашки, **Комисија је закључила да она својим прилазом и тематским одређењем представља новину у истраживањима у овој области.** Главни допринос докторске дисертације се огледа у прецизирању методологије која се може користити за процес анализе утицаја недостајућих података при развоју предиктивних модела и утврђивању да импутација података путем вештачких неуралних мрежа омогућава остваривање бољег предиктивног модела процене метаболичког синдрома у односу на приступ када се недостајући подаци уклањају из полазног скупа.

4. Недостаци дисертације и њихов утицај на резултат истраживања

Докторска дисертација **нема недостатака** који би утицали на коначан резултат истраживања.

X ПРЕДЛОГ:

На основу укупне оцене дисертације кандидата **Дуње Врбаши** под називом **Примена машинског учења у проблему недостајућих података при развоју предиктивних модела** комисија предлаже:

- да се докторска дисертација прихвати, а кандидату одобри одбрана

ПОТПИСИ ЧЛАНОВА КОМИСИЈЕ

Иветић др Драган, редовни професор

Протић др Јелица, редовни професор

Стокић др Едита, редовни професор

Сладић др Горан, ванредни професор

Купусинац др Александар, ванредни професор

Дорословачки др Ксенија, ванредни професор