

УНИВЕРЗИТЕТ У БЕОГРАДУ
ФАКУЛТЕТ ОРГАНИЗАЦИОНИХ НАУКА

НАУЧНО-НАСТАВНОМ ВЕЋУ

Предмет: Реферат о урађеној докторској дисертацији кандидата Сандра Радовановића

Одлуком бр. 05-01 3/82-2 од 23.09.2020. године, именовани смо за чланове Комисије за преглед, оцену и одбрану докторске дисертације кандидата Сандра Радовановића под насловом

Интегрисање доменског знања у развој модела откривања законитости у подацима

После прегледа достављене Дисертације и других пратећих материјала и разговора са Кандидатом Сандром Радовановићем, Комисија је сачинила следећи

РЕФЕРАТ

1. УВОД

1.1. Хронологија одобравања и израде дисертације

Сандро Радовановић је процес израде докторске дисертације започео пријавом теме приступног рада 24.12.2015. године (број одлуке 04-03-16/18). Комисија за оцену научне заснованости пријављене докторске дисертације је формирана 30.12.2015. године бр. 05-01 3/4-7, а приступни рад је одбранио 21.06.2016. године. Одлука о усвајању извештаја Комисије о научној заснованости пријављене докторске дисертације донета је 13.07.2016. године бр. 05-01 3/92-9. Одлуком Универзитета у Београду од 19.09.2016 године бр. 02 61206-4095/2-16 даје се сагласност на предлог теме докторске дисертације Сандра Радовановића под називом „Интегрисање доменског знања у развој модела откривања законитости у подацима“, а за ментора је именован проф. др Милија Сукновић, редовни професор Факултета организационих наука Универзитета у Београду. Ментор је известио Наставно-научно веће Факултета организационих наука да је кандидат Сандро Радовановић завршио израду докторске дисертације, а Наставно-научно веће Факултета организационих наука је именovalo Комисију за оцену завршене докторске дисертације 23.09.2020. године бр. 05-01 3/82-2.

1.2. Научна област дисертације

Дисертација кандидата Сандра Радовановића представља предлог нових методологија за укључивање доменског знања у процес развоја модела откривања законитости у подацима. Односно, централни проблем докторске дисертације је процес извлачења доменског знања о проблему који се решава и његово укључивање у процес учења

модела откривања законитости у подацима, тј. модела одлучивања у виду припреме података и у виду прилагођавања алгоритма.

Докторска дисертација припада области техничких наука и подручју организационих наука. Ужа научна област којом се бави дисертација је Моделирање пословних система и пословно одлучивање. Ментор дисертације је проф. др Милија Сукновић, редовни професор Факултета Организационих Наука, Универзитета у Београду.

1.3. Биографски подаци о кандидату

Сандро Радовановић је рођен 26.07.1989. године у Сарајеву. Основну школу и гимназију општег смера завршио је у Београду са одличним успехом. Факултет организационих наука Универзитета у Београду, студијски програм Информациони системи и технологије, уписао је школске 2008/2009 године. Дипломирао је 2012. године са просечном оценом 9,43 (девет, 43/100) одбраном дипломског рада под називом „Развој система за подршку одлучивању у спорту“. Током студирања више пута је био награђиван за постигнуте академске резултате. Дипломске (мастер) студије на одсеку за Информационе системе и технологије, модул Пословна интелигенција уписао је 2013. године на Факултету организационих наука – Универзитет у Београду. Мастер студије завршио је са просечном оценом 10,00 (десет, 0/100) одбраном мастер рада „Решавање проблема кластеровања генских експресија применом проширеног система мета-учења“.

Практично искуство стиче у Управи за заједничке послове републичких органа, сектор за Информационо-комуникационе технологије, одељење за Информациону безбедност на радном месту Инжењер за техничку заштиту система где је радио од октобра 2012. године до октобра 2017. године. Током рада у Управи за заједничке послове републичких органа обављао је послове програмирања корисничких апликација развоја информатичких система.

Кандидат Сандро Радовановић академско искуство стиче као сарадник у настави ван радног односа при Центру за пословно одлучивање на Универзитету у Београду – Факултет организационих наука током школске 2014/2015 године. У децембру 2017. године постаје сарадник у настави при Центру за пословно одлучивање на Универзитету у Београду – Факултет организационих наука. У децембру 2019. године постаје асистент при Центру за пословно одлучивање на Универзитету у Београду – Факултет организационих наука. Ангажован је на предметима Теорија одлучивања, Пословна интелигенција, Системи за подршку одлучивању и Машинско учење на основним академским студијама и на предметима Системи пословне интелигенције, Складишта података, Откривање законитости у подацима и Развој алгоритама машинског учења на мастер академским студијама. Педагошки рад кандидата се може оценити као успешан са просечном оценом изнад 4,5 (на скали 1-5), што потврђују резултати анкете које Факултет спроводи код студената. Такође, кандидат Сандро Радовановић је два пута је награђиван као најбоље оцењени сарадник у настави, и то у зимском семестру 2018/19 и летњем семестру 2018/19 школске године.

Запослење

- 2012 - 2017: Инжењер за техничку заштиту система, Управа за заједничке послове републичких органа.
- 2017 - 2017: шеф Одсека за пројектовање и имплементацију, Канцеларију за информационе технологије и електронску управу.
- 2014 - 2015: Сарадник у настави ван радног односа, Универзитет у Београду, Факултет организационих наука, Катедра за Организацију пословних система.
- 2017 - 2019: Сарадник у настави, Универзитет у Београду, Факултет организационих наука, Катедра за Организацију пословних система.
- 2019 - тренутно: Асистент, Универзитет у Београду, Факултет организационих наука, Катедра за Организацију пословних система.

Континуирана едукација

2013: Certified RapidMiner "Analyst"

Учешће у пројектима

1. Моделирање доношења одлука у комплексним социо-техничким окружењима, научно-истраживачки пројекат Канцеларије за поморска истраживања Сједињених америчких држава, 2018-2021. Број пројекта: ONR/ONR Global Grant N62909-19-1-2008.
2. Инфраструктура за технолошки подржано учење у Србији, научно-истраживачки пројекат у оквиру програма технолошког развоја Министарства науке Републике Србије, Београд, 2018-2020. Број пројекта ИИИ: 47003
3. Information Security Services Education in Serbia, Еразмус+ KA2 пројекат, 2018-2020. Број пројекта: 586474-EPP-1-2017-1-RS-EPPKA2-CBHE-JP
4. Унапређење предмета везаних за науку о подацима и одлучивање на ФОНу, пројекат Министарства просвете, науке и технолошког развоја у оквиру програма Развоја високог образовања у школској 2018/2019 години. Број пројекта: 451-02-02724/2018-06
5. Предвиђање будућих стања пацијената: Развој и примена брзих, ефективних и интерпретабилних алгоритама за здравство, научно-истраживачки пројекат Швајцарске националне научне фондације, 2014-2016. Број пројекта: IZ3Z0_152415
6. Експлоративна анализа скијашких повреда из болничких отпусних листа, пројекат размене Словенија – Србија, пројекат билатералне сарадње Словеније и Србије, 2014-2015. Број пројекта: BI-RS/14-15-027

Наставно искуство

Основне студије: Теорија одлучивања, Пословна интелигенција, Системи за подршку одлучивању и Машинско учење.

Мастер студије: Системи пословне интелигенције, Складишта података, Машинско учење и Развој алгоритама машинског учења.

2. ОПИС ДИСЕРТАЦИЈЕ

2.1. Садржај дисертације

Докторска дисертација кандидата Сандра Радовановића, под насловом „Интегрисање доменског знања у развој модела откривања законитости у подацима” написана је на 198 страна, груписаних у 6 поглавља, као и дела који се односи на коришћену литературу.

Садржај

1. УВОД.....	1
2. ОСНОВНИ ПОЈМОВИ.....	10
2.1. Откривање законитости у подацима.....	10
2.1.1. Проблеми ОЗП.....	10
2.1.2. Методологија ОЗП	14
2.2. Знање у ОЗП.....	17
2.2.1. Доменско знање	18
2.2.2. Особине знања у ОЗП	19
2.2.3. Репрезентације знања у ОЗП.....	20
2.2.3.1. Симболичке репрезентације	21
2.2.3.2. Просторне репрезентације	23
2.2.3.3. Репрезентација знања преко веза	24
2.2.3.4. Процедуралне репрезентације	25
3. ПРЕГЛЕД СТАЊА У ОБЛАСТИ	26
3.1. Улога доменског знања у процесу ОЗП	26
3.2. Укључивање доменског знања у процес ОЗП	31
3.2.1. Конзистентност доменског знања.....	31
3.2.2. Генерализација доменског знања	32
3.2.3. Конвергенција доменског знања.....	34
3.3. Основни примери укључивања знања у моделе ОЗП	36
3.4. Начини укључивања доменског знања у процес ОЗП	43
3.4.1. Припрема случајева за учење модела ОЗП.....	43
3.4.2. Иницијализација хипотезе или простора хипотеза.....	44
3.4.3. Мењање циља претраге	45
3.4.4. Обогаћивање претраге	46
3.5. Укључивање доменског знања према CRISP-DM методологији.....	47
3.5.1. Разумевање пословног проблема.....	47
3.5.2. Разумевање података.....	48
3.5.3. Припрема података	49

3.5.3.1.	Извођење нових атрибута	49
3.5.3.2.	Избор атрибута.....	53
3.5.3.3.	Вештачки случајеви	61
3.5.4.	Моделовање решења	68
3.5.4.1.	Промена функције циља.....	69
3.5.4.2.	Додавање ограничења.....	71
3.5.4.3.	Развој алгоритма претраге.....	72
3.5.5.	Евалуација решења.....	75
3.5.6.	Примена.....	78
4.	МЕТОДОЛОГИЈА ИСТРАЖИВАЊА	79
4.1.	Предмет истраживања	79
4.2.	Циљеви истраживања.....	80
4.3.	Хипотезе истраживања.....	81
4.4.	Методе истраживања.....	83
4.5.	Подаци	85
4.5.1.	Подаци из области здравства	85
4.5.1.1.	Опис података	85
4.5.1.2.	Софтвер за клиничку класификацију.....	89
4.5.2.	Подаци из области скијашких повреда	90
4.5.2.1.	Опис података	91
4.5.2.2.	Доменско знање у области скијашких повреда.....	94
4.6.	Поставке експеримената	98
5.	РЕАЛИЗАЦИЈА ИСТРАЖИВАЊА	101
5.1.	Припрема података.....	102
5.1.1.	Проширивање скупа атрибута користећи доменско знање.....	102
5.1.1.1.	Експерименти на подацима из области здравства	105
5.1.1.2.	Експерименти на подацима из области скијашких повреда.....	111
5.1.1.3.	Дискусија	119
5.1.2.	Избор атрибута	122
5.1.2.1.	Експерименти на подацима из области здравства	129
5.1.2.2.	Експерименти на подацима из области скијашких повреда.....	134
5.1.2.3.	Дискусија	140
5.2.	Укључивање доменског знања у процес моделовања.....	142
5.2.1.	Развој хијерархијске регуларизације модела ОЗП	143
5.2.1.1.	Експерименти на подацима из области здравства	149
5.2.1.2.	Експерименти на подацима из области скијашких повреда.....	153
5.2.1.3.	Дискусија	157

5.2.2.	Развој ансамбл алгоритма користећи извор доменског знања.....	160
5.2.2.1.	Експерименти на подацима из области здравства	167
5.2.2.2.	Експерименти на подацима из области скијашких повреда.....	171
5.2.2.3.	Дискусија	175
5.3.	Укључивање доменске функције циља	176
5.3.1.	Примена супарничког учења у области здравства.....	180
5.3.2.	Дискусија.....	186
6.	ЗАКЉУЧАК.....	188
	ЛИТЕРАТУРА.....	198

2.2. Кратак приказ појединачних поглавља

У уводном поглављу дат је историјски опис развоја области откривања законитости у подацима и њихова веза са теоријом одлучивања и системима за подршку одлучивању. Посебан осврт је дат на опште друштвено-технолошке околности процеса откривања законитости у подацима које су довеле до предлога теме докторске дисертације. Описана је мотивација аутора за рад у датој области, као и основна идеја рада.

Друго поглавље описује основне појмове који се користе у даљој разради докторске дисертације. Описан је појам, проблеми и методологија откривања законитости у подацима (енгл. *data mining*, у даљем тексту ОЗП). Такође, у овом поглављу је дат опис знања у ОЗП са посебном пажњом на улогу доменског знања у процесу ОЗП, те особинама које знање треба да поседује и начинима репрезентације знања како би се знање било корисно за процес ОЗП.

Преглед стања у области је приказан у трећем поглављу. Представљена је улога доменског знања у процесу ОЗП након чега су изведене особине доменског знања. Односно, описана је конзистентност, генерализација и конвергенција доменског знања за проблем који се решава. Приказан је општи модел укључивања доменског знања у процес ОЗП у виду математичког модела где је могуће укључити доменско знање у две фазе – фази припреме података и у фази учења модела ОЗП (било у виду алгоритма или у виду дефинисања функције циља). Начини укључивања доменског знања су детаљно пописани и груписани по фазама CRISP-DM методологије.

Четврто поглавље описује методологију истраживања. Представљени су предмет, циљеви, хипотезе и методе истраживања. Као генерални циљ истраживања постављен је развој методологије за укључивање доменског знања у поступак развоја модела ОЗП који је разложен на три посебна циља. То су примена екстракције атрибута из других извора знања, развој метода за избор атрибута користећи доменско знање и развој методологије за укључивање доменског знања у алгоритме откривања законитости у подацима. Након описа хипотеза истраживања описани су подаци над којима су спроведена истраживања, као и извори знања који су коришћени за обogaћивање података. Потом је описана поставка истраживања са свим пропратним детаљима неопходним за поновну репродукцију истраживања.

Реализација истраживања, која су подељена на три веће целине, је приказана у петом поглављу. Прва целина се односи на припрему података. Приказане су методологије за проширивање скупа података користећи доменско знање. Прво је скуп атрибута проширен за атрибуте који долазе из других извора знања, а потом је приказана методологија за избор атрибута која се ослања на извор доменског знања. Приказана методологија представља унапређење постојећих метода за избор атрибута зато што укључује извор доменског знања као додатну информацију у процесу прављења модела откривања законитости у подацима. Као резултат добијају се нови изведени атрибути који су разумљиви доносиоцу одлуке, а имају довољну подршку у подацима и добро генерализују проблем који се решава. Друга целина се односи на укључивање доменског знања у процес моделовања. Развијене су две методологије које унапређују мере тачности модела откривања законитости у подацима. Наиме, развијена је хијерархијска регуларизација која кажњава комплексност модела. Допринос ове методологије је у истовременом прављењу и избору атрибута која долази до оптималног решења за задати проблем. Поред ове методологије, приказана је и методологија развоја ансамбл алгоритма која користи извор доменског знања за процес учења модела откривања законитости у подацима. Предложена методологија се ослања на Кондорсетову Теорију пороте и задовољавајући све предуслове наведене теорије постиже боље мере тачности на два показана примера. Трећа целина се односи на укључивање доменске функције циља у процес развоја модела откривања законитости у подацима. Приказана је употреба супарничког учења која решава истовремено две функције циља, од којих једна долази из домена употребе. Сва истраживања су исцрпно евалуирана поредећи се са методама које представљају методе решавања истоветних проблема и у научним и у практичним применама.

На крају, у шестом поглављу, изведени су закључци добијени након реализације експеримената, критички су приказане предности и мане појединачних приступа, као и правци даљег истраживања.

3. ОЦЕНА ДИСЕРТАЦИЈЕ

3.1. Савременост и оригиналност

Тема докторске дисертације је савремена и, како у научном тако и у практичном смислу веома атрактивна. Истраживање и развој методологија за укључивање доменског знања у процес ОЗП има велики значај за унапређивање тачности модела ОЗП и модела одлучивања, те приближавање области управљања знањем и ОЗП.

У докторској дисертацији је предложени потпуно нови поступци за укључивање доменског знања у процес ОЗП, који приступају проблему на оригиналан начин и као такви представљају јединствен допринос научној заједници.

3.2. Осврт на референтну и коришћену литературу

Кандидат је у докторској дисертацији користио изузетно релевантну и обимну домаћу и међународну литературу. Коришћене су публикације еминентних научних удружења у области система за подршку одлучивању, информационих система и технологија, операционих истраживања и машинског учења попут ACM-а и IEEE-а. Велики број коришћених извора је публикован на водећим конференцијама и часописима из

поменутих области. Поред тога коришћена је обимна и актуелна литература коју су публиковали водећи светски издавачи (*Springer, Elsevier, Wiley, Prentice-Hall, Taylor & Francis*).

У изради докторске дисертације кандидат је користио 146 извора релевантних за обраду ове теме.

3.3. Опис и адекватност примењених научних метода

Током израде докторске дисертације коришћене су следеће научне методе:

Почетни део докторске дисертације који се односи на увод у проблем, опис релевантних појмова и преглед литературе, односно поглавља Увод, Основни појмови и Преглед стања у области користи *методе дескрипције* за описивање појава и процеса од интереса, затим *методу анализе* рашчлањивањем сложених појмова, судова и закључака на једноставније саставне делове и елементе, *методу синтезе* кроз уопштавање једноставних судова и доласком до систематизованог знања, као и *методу компилације* критичким разматрањем опажања, ставова и закључака научно-истраживачких радова.

Приликом развоја методологија приказани су математички модели са описаном процедуром оптимизације (коришћене су научне методе *математичка метода* и *метода моделирања*). Описане методологије су евалуиране коришћењем стандардних и модерних метода евалуације класификационих модела. Резултати су упоређени стандардним методама евалуације у области откривања законитости у подацима, али и традиционалним статистичким тестовима. Утврђене су сличности и различитости у добијеним резултатима и показано је када ће предложене методологије резултовати бољим у мерама тачности.

Начин на који је коришћена литература, њени извори и критички осврт јасно указују на адекватан истраживачки приступ. Реализација истраживања и експериментална евалуација је постављена адекватно и у складу са стандардима за области откривања законитости у подацима.

На основу анализе докторске дисертације, може се закључити да су примењене научне методе и технике одговарајуће по свом значају, структури, теми дисертације и спроведеном истраживању.

3.4. Применљивост остварених резултата

Предложене методологије су успешно проверене емпиријски, на примерима из здравства и области скијашких повреда. Резултати су показали да хипотеза: *укључивање доменског знања у развој модела откривања законитости у подацима ће резултовати побољшањем његових перформанси* не може бити одбачена. Предложене методологије за проширивање скупа атрибута користећи доменско знање, те обе методологије за избор атрибута и предложени ансамбл алгоритама су показали боље мере тачности у односу на методе које су коришћене за поређење. Све предложене методологије су развијене тако да се могу примењивати и у другим применама, уколико је извор доменског знања сличног облика.

3.5. Оцена достигнутих способности кандидата за самостални научни рад

Кандидат је у досадашњем научно-истраживачком раду показао изузетну способност да истражи и критички анализира релевантну литературу, идентификује истраживачка питања и на основу њих постави и спроведе истраживање. Истакао се учешћем у водећим националним и међународним конференцијама које су као фокус имале област дисертације. У току досадашњем научно-истраживачког рада и током рада на дисертацији објавио је као аутор или коаутор 62 рада у водећим међународним и националним часописима, као и зборницима радова националних и међународних конференција. Као истраживач, учествовао је, или и даље учествује, у једном домаћем научно-истраживачком пројекту, као и у три међународна пројекта. Треба споменути да је кандидат, током докторских студија, усмеравао и надгледао научно-истраживачки рад млађих колега са којима је у сарадњи остварио високе резултате.

Узевши у обзир целокупно залагање и остварене резултате у научно истраживачком раду, закључујемо да је кандидат способан да се у потпуности самостално бави научно-истраживачким радом.

4. ОСТВАРЕН НАУЧНИ ДОПРИНОС

4.1. Приказ остварених научних доприноса

Област укључивања доменског знања у процес развоја модела откривања законитости у подацима представља област која је интересантна за више области поред откривања законитости у подацима. Научни допринос се огледа у систематизовању литературе и детаљним прегледом стања у области. Међутим, најзначајнији допринос јесу у методологији. Развијене су оригиналне методологије за избор атрибута која се заснива на похлепној хеуристици, као и на приступу регуларизације модела откривања законитости у подацима. Затим, развијена је и оригинална методологија за ансамбл алгорита заснован на доменском знању. Наведене методологије укључују извор доменског знања у процес развоја модела откривања законитости у подацима било у фази припреме података или у фази укључивања доменског знања у процес моделовања.

Систематизовано, научни допринос докторске дисертације се огледа у следећем:

- Преглед и систематизација савремених метода за укључивање доменског знања у моделе откривања законитости у подацима.
- Примена проширивања скупа атрибута користећи доменско знање на реалним скуповима података, детаљна анализа резултата и поређење са традиционалним приступом.
- Развој методологије за избор атрибута која се заснива на доменском знању користећи похлепну хеуристику.
- Развој методологије за избор атрибута која се заснива на доменском знању користећи регуларизацију модела откривања законитости у подацима.
- Примена избора атрибута која се заснива на доменском знању (и као фаза припреме података и у фази моделовања) на реалним скуповима података, детаљна анализа резултата и поређење са другим методама из литературе.
- Развој методологије за ансамбл алгорита која се заснива на извору доменског знања.

- Примена методологије за ансамбл алгоритам која се заснива на извору доменског знања на реалним скуповима података, детаљна анализа резултата и поређење са другим методама из литературе.
- Примена супарничког учења као методе за укључивање доменске функције циља на реалном скупу података.

4.2. Критичка анализа резултата истраживања

Реализована истраживања и добијени резултати истраживања представљају целовит преглед области укључивања доменског знања у развој модела откривања законитости у подацима, нов предлог методологије за развој хијерархијске регуларизационе технике модела ОЗП која се заснива на доменског знању, те нове методологије за развој ансамбл модела ОЗП који користи извор доменског знања. Методологије су експериментално проверене реализацијом предложених методологија на два примера и поређењем са методологијама које се користе у практичним применама. Показано је да предложене методологије доприносе побољшању предиктивних перформанси модела ОЗП.

4.3. Верификација научних доприноса

Области научног интересовања Сандра Радовановића су теорија одлучивања, машинско учење, откривање законитости у подацима (енгл. *Data Mining*), системи пословне интелигенције, системи за подршку одлучивања и складишта података. Током свог научно истраживачког рада објавио је велики број научних радова и то у следећим категоријама: M13 – један рад, M21 – осам радова, M22 – два рада, M23 – три рада, M33 – двадесет девет радова, M34 - десет радова, M51 – три рада, M52 - један рад и M63 – два рада.

У наставку је дат је одабрани подскуп радова у којима су директно или индиректно описани кључни доприноси из теме ове дисертације разврстани по категоријама.

Категорија M21:

1. Jovanović, M., **Radovanović, S.**, Vukićević, M., Van Poucke, S., & Delibašić, B. (2016). Building interpretable predictive models for pediatric hospital readmission using tree-lasso logistic regression. *Artificial Intelligence in Medicine*, 72(9) (pp. 12-21). Elsevier Publishing. DOI: 10.1016/j.artmed.2016.07.003. IF(2015)=2.142

Категорија M22:

1. **Radovanovic, S.**, Delibasic, B., Suknovic, M., & Matovic, D. (2019). Where will the next ski injury occur? A system for visual and predictive analytics of ski injuries. *Operational Research*, 1-20. DOI: 10.1007/s12351-018-00449-x. IF(2018)=1.816

Категорија M23:

1. **Radovanović, S.**, Delibašić, B., Jovanović, M., Vukićević, M., Suknović, M. (2019) A Framework for Integrating Domain Knowledge in Logistic Regression with Application to Hospital Readmission Prediction. *International Journal on Artificial Intelligence Tools*, 28(6), 19 pages. DOI: 10.1142/S0218213019600066. IF(2018)=0.849

Категорија M13:

1. Delibašić B., **Radovanović S.**, Jovanović M.Z., Suknović M. (2020) Improving Decision-Making in Ski Resorts by Analysing Ski Lift Transportation—A Review. In: Mladenović N., Sifaleras A., Kuzmanović M. (eds) *Advances in Operational Research in the Balkans*. Springer Proceedings in Business and Economics. Springer, Cham. DOI: 10.1007/978-3-030-21990-1_16

Категорија M33:

1. **Radovanovic, S.**, Vukicevic, M., Kovacevic, A., Stiglic, G., Obradovic, Z. (2015). Domain knowledge Based Hierarchical Feature Selection for 30-Day Hospital Readmission Prediction. In *Proceedings of the 15th Conference on Artificial Intelligence in Medicine, AIME 2015* (pp. 96-100). Pavia, Italy: Springer International Publishing. DOI:10.1007/978-3-319-19551-3_11
2. Vukicevic, M., **Radovanovic, S.**, Kovacevic, A., Stiglic, G., & Obradovic, Z. (2015). Improving Hospital Readmission Prediction Using Domain Knowledge Based Virtual Examples. In *Knowledge Management in Organizations* (pp. 695-706). Springer International Publishing. DOI: 10.1007/978-3-319-21009-4_51
3. Vukicevic, M., **Radovanovic, S.**, Stiglic, G., Delibasic, B., Van Poucke, S., Obradovic, Z. (2016). A Data and Knowledge Driven Randomization Technique for Privacy-Preserving Data Enrichment in Hospital Readmission Prediction. In *Proceedings of 5th Workshop on Data Mining in Medicine and Healthcare – SDM DMMH 2016* (pp. 10-18). Miami, Florida, USA.
4. **Radovanović, S.**, Delibašić, B., Jovanović, M., Vukićević, M., Suknović, M. (2018) Framework for integration of domain knowledge into logistic regression. In *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics (WIMS '18)*, (Article No. 24). Rajendra Akerkar, Mirjana Ivanović, Sang-Wook Kim, Yannis Manolopoulos, Riccardo Rosati, Miloš Savić, Costin Badica, and Miloš Radovanović (Eds.). ACM, New York, NY, USA, Article 24, 8 pages. DOI: 10.1145/3227609.3227653

5. ЗАКЉУЧАК И ПРЕДЛОГ

5.1. Кратак осврт на дисертацију у целини, на научне доприносе конкретној научној области и методологији, на примену у пракси и на способности кандидата

Докторска дисертација се бави начинима и метода укључивања доменског знања у процес учења модела откривања законитости у подацима. Како доменско знање и даље има велику улогу у процесу доношења одлука, способност да се правила и законитости која се налазе у подацима интегришу са доменским знањем је од великог значаја.

Циљ докторске дисертације је био да се развије методологија за укључивање доменско знања у поступак развоја модела откривања законитости у подацима идентификовањем кључних места у процесу откривања законитости у подацима подложним утицајем доменског знања. Представљен је утицај доменског знања кроз општи математички модел откривања законитости у подацима, као и кроз фазе процеса откривања законитости у подацима. Наведени математички модел представља преглед стања у обрађеној области чиме је дата потпуна слика академских активности у посматраној области. На основу релевантне литературе и критичког осврта идентификован је простор за унапређење кроз развој нових методологија.

Представљањем кроз општи математички модел откривања законитости у подацима, кандидат је идентификовао три места где је било могуће укључити доменско знање. То су екстракција атрибута, избор атрибута и развој алгоритама откривања законитости у подацима заснованих на доменском знању. Кандидат је у докторској дисертацији развио три оригиналне методологије за укључивање доменског знања у процес развоја модела откривања законитости у подацима. Кандидат је развио методу засновану на похлепној хеуристици за избор атрибута која користи извод доменског знања за избор атрибута који има довољну подршку у подацима. Затим, како претходна поменута метода нема гаранцију оптималности, развијена је регуларизациона техника која даје гаранције оптималног решења приликом избора атрибута. На крају, развијена је оригинални ансамбл алгоритама који користи Кондорсетову Теорију пороте за прављење већег броја модела откривања законитости у подацима. Поред наведених развијених оригиналних методологија кандидат је применио методу супарничког учења за постизање Парето оптималног решења и функције циља модела откривања законитости у подацима и доменске функције циља. Све наведене методологије су научно засноване и објашњене како кроз математичку нотацију, тако и кроз примере.

Предложене методологије су експериментално проведене на реалним скуповима података из области здравства и из области скијашких повреда. Низом извршених експеримената испитане су мере тачности развијених метода и упоређене са релевантним метода применом методе поређења и статистичком анализом. На основу добијених резултата долази се до закључка да се постижу унапређења у мерама тачности.

Имајући у виду постигнуте научно-истраживачке резултате кандидата, претходно поменуте предложене методологије укључивања доменског знања у процес развоја откривања законитости у подацима, њихову експерименталну проверу, као и актуелност и специфичности обрађене теме, констатујемо да докторска дисертација Сандра Радовановића задовољава највише критеријуме и показује способност кандидата за самосталан научно-истраживачки рад.

5.2. Предлог Комисије Научно-наставном већу

Предлажемо да се приложена докторска дисертација под називом „Интегрисање доменског знања у развој модела откривања законитости у подацима“ кандидата Сандра Радовановића прихвати, изложи на увид у јавности и упути на коначно усвајање већу техничких наука Универзитета у Београду.

ЧЛАНОВИ КОМИСИЈЕ

Проф. др Милија Сукновић, редовни професор
Универзитет у Београду, Факултет Организационих Наука

Проф. др Борис Делибашић, редовни професор
Универзитет у Београду, Факултет Организационих Наука

Проф. др Зоран Обрадовић, редовни професор
Универзитет Темпл – одељење за рачунарске и информационе науке