

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

I ПОДАЦИ О КОМИСИЈИ
<p>1. Датум и орган који је именовao комисију 22.10.2015, Сенат Универзитета у Новом Саду</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>1. Проф. др Андреја Тепавчевић, редовни професор, уже научна област Алгебра и логика, 1.12.2003, Природно-математички факултет у Новом Саду, председник</p> <p>2. Проф. др Зорана Лужанин, редовни професор, уже научна област Нумеричка математика, 12.11.2007, Природно-математички факултет у Новом Саду, члан комисије, ментор</p> <p>3. Проф. др Наташа Крејић, редовни професор, уже научна област Нумеричка математика, 14.6.2004, Природно-математички факултет у Новом Саду, члан</p> <p>4. Проф. др Вера Грујић, редовни професор у пензији, уже научна област Социјална медицина, 19.12.1999, Медицински факултет у Новом Саду, члан</p> <p>5. Проф. др Светлана Квргић, ванредни професор, уже научна област Социјална медицина, 24.5.2007, Медицински факултет у Новом Саду, члан</p>
II ПОДАЦИ О КАНДИДАТУ
<p>1. Име, име једног родитеља, презиме: Наташа (Раде) Драгнић</p> <p>2. Датум рођења, општина, држава: 19.05.1973, Сомбор, Србија</p> <p>3. Назив факултета, назив студијског програма дипломских академских студија – мастер и стечени стручни назив Природно-математички факултет Нови Сад, смер дипломирани математичар</p> <p>4. Година уписа на докторске студије и назив студијског програма докторских студија Кандидаткиња није похађала докторске студије</p> <p>5. Назив факултета, назив магистарске тезе, научна област и датум одбране: Природно-математички факултет Нови Сад, <i>Примена логистичког регресионог модела на одређивање повезаности гојазности са потенцијалним факторима ризика</i>, Математика, 27.6.2008.</p> <p>6. Научна област из које је стечено академско звање магистра наука: Математика</p>
III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:
<p>Конструкција и анализа кластер алгорита са применом у дефинисању бихејвиоралних фактора ризика у популацији одраслог становништва Србије.</p>

IV ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

Докторска дисертација је написана на српском језику и садржи 154 стране. Дисертација обухвата 10 поглавља, 25 табела, 3 слике, 2 графикона и 3 прилога. Списак литературе садржи 272 референце.

Текст докторске дисертације је приказан по следећим поглављима:

1. Увод
 2. Примена нумеричких метода оптимизације у решавању проблема кластеровања
 3. Анализа кластер алгоритама
 4. Циљеви и хипотезе истраживања
 5. Кластеровање великих скупова података. Резултати
 6. Бихејвиорални фактори ризика
 7. Бихејвиорални фактори ризика и кластер анализа. Резултати
 8. Дискусија
 9. Закључци
 10. Литература
- Прилози

У уводном разматрању дефинисан је појам кластер анализе, као и самог проблема кластеровања. Уведене су основне дефиниције из неглатке анализе и теорије вероватноће и статистике, а које се касније користе у дисертацији. На крају поглавља је уведен и објашњен појам сложености алгоритама (просторне и временске).

Проблем одређивања кластера математички се моделира као проблем нелинеарне оптимизације са ограничењима, који се у највећем броју случајева, може решити само применом метода нумеричке оптимизације. У другом поглављу је приказан неглатки оптимизациони приступ у кластер анализи. Описани су алгоритам k -средина, глобални алгоритам k -средина, као и кластер алгоритам заснован на неглаткој оптимизацији. Две карактеристике датог скупа података могу значајно утицати на резултате кластеровања: број објеката и број променљивих, те је у овом делу дисертације описан поступак за редуковање броја објеката у скупу, без губитка информација.

У трећем поглављу је описана проблематика кластеровања великих скупова података са категоријалним и комбинованим типовима обележја, са освртом на два алгоритама: алгоритам k -модуса и његове модификације, као и двостепени кластер алгоритам. На крају овог поглавља описане су методе за евалуацију валидности резултата кластер анализе.

У четвртом поглављу су дефинисани циљеви и хипотезе истраживања, описана је методологија и коришћене статистичке методе у дисертацији. Први део методологије садржи опис података коришћених у медицинском делу дисертације (*Истраживање здравља становника Србије 2006. године*), као и делу резултата за комбиноване типове обележја. Овај део се састоји из описа узорка, упитника коришћених у истраживању, као и обележја коришћених у кластер анализи. Други део методологије представља опис базе података *Mushrooms*, коришћене у делу резултата за велике скупове података са категоријалним променљивима.

Пето поглавље представља резултате за кластеровање великих скупова података са категоријалним и комбинованим типовима података. Резултати су анализирани на целом скупу података, а затим на простим случајним узорцима. Описан је модификовани поступак кластеровања за велике скупове података где најважнији корак поступка представља коришћење простих случајних узорака одређене кардиналности, уместо целог скупа података.

У шестом поглављу су дефинисани и описани бихејвиорални фактори ризика: недовољна физичка активност, пушење, злоупотреба алкохола и неправилна исхрана, као и утицај комбинованог деловања два или више фактора ризика на здравље појединца. Метод кластер анализе омогућава холистички приступ у идентификацији релевантних циљних група за анализу фактора ризика и предвиђање одређених интервентних мера.

На који начин се кластерују бихејвиорални фактори ризика у популацији одраслог становништва приказано је у седмом поглављу. Мултиваријантна анализа је примењена у циљу утврђивања повезаности социодемографских карактеристика одраслог становништва и добијених кластера бихејвиоралних фактора ризика.

Дискусија се састоји из три дела. У првом делу су анализирани добијени резултати са освртом на предности, односно недостатке предложеног приступа у кластеровању великих скупова података. У другом делу су анализирани добијени кластери бихејвиоралних фактора ризика са специфичним социодемографским карактеристикама, а добијени резултати су упоређени са резултатима других истраживања. Трећи део овог поглавља представља детаљну анализу ограничења истраживања.

Закључци су приказани у деветом поглављу. Након референтно наведене литературе, у прилогу су дати упитници коришћени у *Истраживању здравља становништва Србије 2006. године* (Упитник за домаћинство, Упитник за одрасле, Упитник за самопопуњавање).

V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

Кластер анализа има дугу историју и мада се примењује у многим областима и даље остају значајни изазови. Један од главних изазова у кластер анализи је рад са великим скуповима података са категоријалним и комбинованим (нумерички и категоријални) типовима променљивих. Рад са великим бројем инстанци (објеката) и великим бројем димензија (променљивих) може представљати проблем у кластер анализи, због временске сложености. Један од начина решавања овог проблема је редуковање броја инстанци, без губитка информација. Преглед најзначајнијих резултата у литератури везан за ову проблематику је приказан у прва три поглавља.

Научни резултати докторске дисертације су најбоље истакнути у петом, седмом и осмом поглављу. Први циљ дисертације је био упоређивање резултата кластеровања на целом скупу и простим случајним узорцима са категоријалним и комбинованим подацима, за различите величине узорка и различит број кластера. Утврђено је високо слагање резултата кластеровања на узорцима и целом скупу, а изузетак чине узорци величине $0.01m$ (где је m обим основног скупа). Резултати су детаљно приказани у поглављу пет.

Други циљ дисертације је био конструкција ефикасног поступка кластеровања великих скупова података са категоријалним и комбинованим типовима променљивих. Модификовани поступак кластеровања представља комбинацију поступка одабира простих случајних узорака одређене кардиналности и примене одговарајућег алгорита кластеровања на овим узорцима. Коришћење простих случајних узорака обезбеђује репрезентативност података, па се резултати добијени на узорцима могу уопштити на цео скуп. Предложени поступак је описан у петом поглављу, а дискутован у осмом поглављу, са освртом на његове предности и недостатке.

Трећи циљ дисертације представља примену кластер анализе у дефинисању кластера бихејвиоралних фактора ризика у популацији одраслог становништва Србије, као и анализу социодемографских карактеристика добијених кластера. Мада је кластер анализа веома распрострањена у социологији, комерцијалном истраживању маркетинга и многим другим областима, још увек није довољно разматрана њена примена у социјалној медицини, епидемиологији, уопште у области јавног здравља. Кластер анализа је у већем броју истраживања из области јавног здравља ограничена на корелацију између два бихејвиорална фактора ризика и не разматрају се кластери формиран на основу мултидимензионалних карактеристика. Значај примене кластер анализе у дефинисању кластера бихејвиоралних фактора ризика је истакнут у поглављу шест, а резултати су приказани у поглављу седам. Добијени резултати су упоређени са резултатима других истраживања, што је приказано у поглављу осам. Информације о социодемографским карактеристикама добијених кластера су корисне у планирању будућих превентивних стратегија, јер указују на циљне категорије становништва које су приоритетне у планирању и спровођењу специфичних превентивних програма и интервенција у будућности. Приликом интерпретације резултата кластеровања веома је важно

указати на постојећа ограничења истраживања, што је детаљно анализирано у поглављу осам (Дискусија).

Представљени резултати су добро илустровани и прецизно тумачени, јасно потврђујући постављене циљеве, односно хипотезе истраживања. У дискусији и закључном разматрању истакнути су најбитнији доприноси овога рада, као и могући правци даљих истраживања на ову тему. Закључци су доследно, сажето и прецизно изведени. У целини посматрано, дисертација је написана систематично и јасно. Изложено решење у контексту одабране проблематике предстваља оригиналан и вредан допринос науци.

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.

Следећи радови су настали на основу резултата истраживања ове докторске дисертације и осталих истраживања у току израде докторске дисертације:

1. Grujić V, Martinov-Cvejic M, Ač Nikolić E, Draganić N, Mijatović Jovanović V, Kvirgić S, Travar S. Association between obesity and socioeconomic factors and lifestyles, *Vojnosanitetski pregled* 2009; 66 (9): 705-710 (M23)
2. Grujić V, Draganić N, Harhaji S, Radić I, Šušnjević S. Overweight and Obesity among Adults in Serbia: Results from the National Health Survey, *Eating and Weight Disorders* 2010; 15(1-2): 34-42 (M23)
3. Grujić V, Draganić N, Kvirgić S, Susnjević S, Grujić J, Travar S. Epidemiology of Hypertension in Serbia: Results of a National Survey. *Journal of Epidemiology* 2012; 22(3): 261-266 (M21)
4. Grujić V, Draganić N, Mijatović Jovanović V, Ukropina S, Harhaji S, Radić I, Kvirgić S. Predictors of overweight and obesity among adults aged 50 years and above: Serbian National Health Survey. *Vojnosanitetski pregled* (прихваћен за штампу, септембар 2015) (M23)

VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА

За велике скупове података са комбинованим подацима коришћена је база *Истраживање здравља становништва Србије*. Примењен је TSCA алгоритам на целом скупу, при чему је коришћењем Бајесовог информационог критеријума (BIC) утврђен оптималан број кластера $k = 5$. Као мера интерне валидности коришћен је *Silhouette* индекс. Даљим коришћењем поступка вишеструког извлачења простих случајних узорака величине $0.01m, 0.03m, 0.05m, 0.1m, 0.3m$ (где је m обим основног скупа) и применом TSCA алгоритма за комбинована обележја на узорцима, утврђен је такође најоптималнији број кластера за $k = 5$. Највећа тачност, то јест слагање резултата кластеровања на узорцима са добијеним кластерима на целом скупу је такође потврђена за овај оптимални број кластера. Изузетак чине узорци величине $0.01m$.

За велике скупове података са категоријалним подацима коришћена је база *Mushrooms*, са већ дефинисаном класификацијом. Упоредивањем различитих алгоритама кластеровања на целом скупу података, утврђени су најбољи резултати за TSCA и Ward-ov алгоритам кластеровања. Најбоље кластерско решење је одређено на основу критеријума екстерне валидности-тачности, то јест поклапања резултата кластеровања са већ постојећом класификацијом на целом скупу. Недостатак Ward-ovog алгоритма за велике скупове података представља његова временска сложеност.

У циљу утврђивања да ли се смањењем улазних података (обима скупа), а самим тим и временске сложености алгоритма, не нарушава структура добијених кластера на целом скупу, Ward-ov алгоритам је даље примењен на категоријалне податке, коришћењем простих случајних узорака (величине $0.01m, 0.03m, 0.05m, 0.1m, 0.3m$, где је m обим основног скупа). Резултати добијени на узорцима су показали високо слагање са постојећом класификацијом на целом скупу. Изузетак чине узорци величине $0.01m$.

Поступак вишеструког извлачења (i_{\max} број извлачења) простих случајних узорака величине $t_u m$ (коэффициент $t_u \in (0,1)$, m обим основног скупа) и примена кластер алгоритма на овим узорцима даје једнако добре резултате (није нарушена структура кластера) као и алгоритам примењен на целом скупу. Предложени модификовани поступак кластеровања великих скупова података се састоји из следећих фаза:

1. Кластеровање на простим случајним узорцима одређене кардиналности
2. Применом одговарајућег критеријума валидности добија се најбоље кластерско решење на i' -том узорку ($i' \leq i_{\max}$)
3. Добijени центри кластера из i' -тог узорка служе за кластеровање остатка скупа.

Применом двостепеног кластер алгоритма издвајају се јасно одвојени кластери у популацији одраслог становништва Србије, са карактеристичним комбинацијама бихејвиоралних фактора ризика: *Без ризичних фактора, Штетна употреба алкохола и друге ризичне навике, Неправилна исхрана и друге ризичне навике, Недовољна физичка активност, Пушачи*. Један од начина процене валидности добијеног кластерског решења обухвата тестирање разлика између кластера на неким релевантним екстерним променљивама, које нису коришћене у поступку кластеровања. Тестирана је разлика између кластера бихејвиоралних фактора ризика у односу на социодемографске карактеристике (екстерне променљиве). Применом мултиномног логистичког регресионог модела, закључујемо да испитаници који нису у браку, лошијег су материјалног стања, нижег образовања и живе у Војводини имају већу шансу за присуство вишеструких бихејвиоралних фактора ризика.

Истраживање даје увид у удруживање бихејвиоралних фактора ризика и коваријате (придружене факторе) за бихејвиоралне факторе ризика код појединаца, а резултати кластер анализе обезбеђују више свеобухватних информација о здравственом стању становништва него појединачни фактори ризика.

VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА

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

Докторска дисертација је коректно и систематично приказала:

-стање у области истраживања (на основу доступне литературе)

-постигнуте резултате истраживања (уз резултате су наведена и јасна и одговарајућа образложења)

-поређење добијених резултата са резултатима других истраживања, као и даље правце истраживања.

На основу свега наведеног, Комисија позитивно оцењује начин приказа и тумачење постигнутих резултата истраживања.

IX КОНАЧНА ОЦЕНА ДОКТОРСKE 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

ДА

2. Да ли дисертација садржи све битне елементе

ДА

3. По чему је дисертација оригиналан допринос науци

У дисертацији је приказан приступ у решавању кластерована великих скупова података са категоријалним, односно комбинованим типовима података, са аспекта нумеричке математике и са аспекта области јавног здравља. Кандидаткиња је анализирала проблеме који се јављају у раду са оваквим подацима и развила поступак чија примена представља основу за решење поменутих проблема. Предложени модификовани поступак кластерована прави компромис између добијене прецизности (незнатно нижа у односу на цео скуп) и постигнуте ефикасности (мања временска сложеност). Неопходна је даља анализа случајева за различит број променљивих укључених у кластер анализу, а оптимална вредност коефицијента $t_u \in (0,1)$ је свакако предмет даљих истраживања.

Анализа података (у медицинском делу дисертације) је спроведена на великом репрезентативном узорку одраслог становништва Србије и представља прву студију која се бави испитивањем бихејвиоралних фактора ризика у овој популацији, применом кластер анализе. Са аспекта јавног здравља, кластер анализа идентификује фиксирани и стабилни кластери у испитиваном скупу података, омогућавајући давање препорука за модификацију постојећих навика повезаних са стилем живота код одраслог становништва, као и дизајнирање будућих анкетних упитника за прикупљање података.

4. Недостаци дисертације и њихов утицај на резултат истраживања

Докторска дисертација не садржи недостатке који би утицали на резултате истраживања.

X ПРЕДЛОГ:

На основу укупне оцене дисертације, Комисија предлаже да се докторска дисертација под називом

Конструкција и анализа кластер алгоритма са применом у дефинисању бихејвиоралних фактора ризика у популацији одраслог становништва Србије

кандидаткиње Наташе Драгнић прихвати, а кандидаткињи одобри одбрана.

ЧЛАНОВИ КОМИСИЈЕ

др Андреја Тепавчевић, редовни професор,
председник

др Зорана Лужанин, редовни професор,
члан, ментор

др Наташа Крејић, редовни професор,
члан

др Вера Грујић, редовни професор у пензији,
члан

др Светлана Квргић, ванредни професор,
члан

НАПОМЕНА: Члан комисије који не жели да потпише извештај јер се не слаже са мишљењем већине чланова комисије, дужан је да унесе у извештај образложење односно разлоге због којих не жели да потпише извештај.