УНИВЕРЗИТЕТ У БЕОГРАДУ

ФАКУЛТЕТ ОРГАНИЗАЦИОНИХ НАУКА

Ана М. Поледица

ЛОГИЧКИ ПРИСТУП МОДЕЛОВАЊУ СЛИЧНОСТИ

докторска дисертација

Београд, 2016.

UNIVERSITY OF BELGRADE

FACULTY OF ORGANIZATIONAL SCIENCES

Ana M. Poledica

A LOGIC-BASED APPROACH TO SIMILARITY MODELING

doctoral dissertation

Belgrade, 2016

МЕНТОР:

____

др Братислав Петровић, редовни професор

Факултет организационих наука, Универзитет у Београду

ЧЛАНОВИ КОМИСИЈЕ:

____

др Милан Мартић, редовни професор

Факултет организационих наука, Универзитет у Београду

____

др Драган Радојевић, научни саветник

Институт „Михајло Пупин“, Универзитет у Београду

Датум одбране дисертације: ___________________

Пре свега се захваљујем свом ментору проф. др Петровићу на поверењу које ми је указао и прилици за остварење академске каријере на Факултету организационих наука. Такође се захваљујем на стручним и пријатељским саветима и подршци у току израде ове докторске дисертације. Захваљујем се и члановима комисије проф. др Радојевићу на инспиративној теорији која је била покретач за ово истраживање, и проф. др Мартићу на поверењу и подршци у току дугогодишњег рада на Факултету.

Захваљујем се колегама са Катедре за управљање системима Ивани Драговић, Александру Ракићевићу, а посебно Павлу Милошевићу на пријатељској подршци, конструктивним саветима и дискусији. Захвална сам и осталим колегама на Факултету на саветима и помоћи.

Хвала драгим пријатељима на стрпљењу и разумевању у току израде докторске дисертације.

Посебно се захваљујем својој породици на безрезервној љубави и подршци на сваком кораку мога пута.

ЛОГИЧКИ ПРИСТУП МОДЕЛОВАЊУ СЛИЧНОСТИ

Резиме: У овој докторској дисертацији уведен је логички приступ моделовању сличности који је заснован на интерполативној Буловој алгебри. За мерење сличности, предложене су нове интерпретабилне логичке мере, параметарске и непараметарске, као и дескриптивни оператор агегације – логичка агрегација. Поред пружања теоријске основе, у овом истраживању посебна пажња је посвећена емпиријској анализи. У сврху валидације дефинисаних мера уведена је логичка класификација заснована на ИБА сличности. За све уведене мере извршена је евалуација и поређење на реалним подацима из домена медицине, где је показано да увођење параметара унапређује резултате класификације. На крају су приказане могућности за конструисање логичких класификатора заснованих на експертским функцијама сличности на проблему предвиђања банкротства предузећа.

Кључне речи: моделовање сличности, интерполативна Булова алгебра, логичке мере сличности, параметарске мере сличности, класификација, класификација заснована на сличности

Научна област: Техничке науке

Ужа научна област: Управљање системима

УДК број: 519.7

A LOGIC-BASED APPROACH TO SIMILARITY MODELING

Abstract: In this doctoral thesis, a logical approach to similarity modeling based on interpolative Boolean algebra is introduced. Novel interpretable logical measures, both nonparametric and parametrized, are introduced for measuring similarity together with logical aggregation as a descriptive aggregation operator. Besides the theoretical background, in this research special attention is devoted to empirical analysis. For validation purposes, logical classification based on IBA similarity is introduced. Defined logical measures are evaluated and compared in the case of medical data, and it is shown that parameterized measures improve classification results. Finally, the benefits of logic-based classifiers with expert similarity functions are presented on the problem of corporate bankruptcy prediction.

Key words: similarity modeling, interpolative Boolean algebra, logic-based similarity measures, parameterized similarity measures, classification, similarity-based classification

Scientific field: Technical sciences

Specific scientific field: System control

UDK number: 519.7

САДРЖАЈ

1. УВОД

1.1. Проблем, предмет и циљ истраживања

Класична логика дуго је представљала формалну основу за проучавање људског расуђивања. Последњих деценија постало је јасно да у реалном/свакодневном животу људско резоновање (*енг.* human reasoning) захтева више него што традиционална дедуктивна логика може да пружи *(Godo & Rodriguez, 2008).*

Класична логика се сматра неодговарајућом за моделовање људског резоновања јер је потребно да се узме у обзир „несавршено" знање у облику неодређености, неизвесности, непотпуности, итд. *(Minsky, 1985; McDermott 1987).* Услед тога, последњих деценија дошло је до експанзије логичких формализама који су се показали погодним за моделовање проблема у чијој је основи људско расуђивање. Ови логички формализми односе се пре свега на вишевредносну и фази логику, вероватносну логику или неку њихову комбинацију *(Godo & Rodriguez, 2008).*

Сличност односно различитост представља један од фундаменталних концепата који се налази у основи целокупног људског расуђивања и искуства *(Deza & Deza, 2014).* Мерење сличности се обично реализује помоћу мера заснованим на геометријским моделима чије метричке аксиоме не морају бити задовољене у практичним применама где се процена сличности врши од стране људи *(Tversky, 1977; Rorissa, 2007).* Док се са једне стране, број мера и њихових примена повећава, са друге стране не постоји одговарајућа теоријска потпора да прецизно објасни како оне мере сличност и које од њих треба изабрати за решавање одређеног проблема *(Janowicz et al., 2011).* Критичка анализа метричких својстава при креирању функција сличности представља полазну мотивацију за истраживање у овом раду. Такође, при мерењу сличности један од најбитнијих проблема је разлика која постоји између идеализованих мера и саме људске перцепције сличности. Снага неког модела сличности треба да се мери и на основу тога колико добро објашњава природу људске оцене сличности *(Rorissa, 2007).* Поред тога, метрички постулати ограничавају доменске експерте при моделовању сличности *(Eckhardt et al., 2009).* Због тога су као алтернатива метричким приступима настали новији правци, међу којима се посебно издвајају мере сличности које у основи имају неку логичку релацију импликације или еквиваленције *(Luukka, 2005; Saastamoinen, 2006; Le Capitaine, 2012).* Ове логичке мере сличности се заснивају на вишевредносним релацијама импликације и еквиваленције. *Zadeh (1971)* је у једном од својих раних радова увео фази еквиваленцију као природан начин да се опише људска перцепција сличности.

Поред тога, увођењем параметара може се уопштити функција сличности и проширити домен њене примене. Да би модел сличности обухватио претходно искуство акумулирано у подацима, може се користити обучавање параметара из података.

Проблем истраживања докторског рада је моделовање сличности помоћу интерполативне Булове алгебре (ИБА) *(Radojević, 2000)* као логичког формализма. ИБА је реално-вредносна реализација Булове логике на јединичном интервалу. Мерење сличности представља природно место за увођење ИБА јер сличност по својој суштини подразумева третирање градације али и услед потребе за што јаснијом интерпретацијом модела сличности.

Предмет докторског рада је наставак истраживања логичког приступа мерењу сличности заснованог на ИБА еквиваленцији и логичкој агрегацији (*Radojević, 2008a*). Поред тога, у раду су разматране могућности параметризације логичких модела сличности који ће се евалуирати на проблему класификације. Сличност се користи као принцип организовања помоћу кога појединци класификују објекте, формирају концепте и заснивају генерализацију *(Tversky, 1977).* У том смислу, различити модели сличности, параметарски и непараметарски, могу бити основа за дефинисање алгоритама за класификацију, кластеровање, селекцију карактеристика, редукцију података, детекцију изузетака, регресију и претраживање.

Значај логичких приступа моделовању сличности се огледа и у томе што пружају формалну основу за проучавање различитих проблема који у својој природи поседују градацију, где важну улогу има начин на који се третира неодређеност али и где је интерпретација резултата важна за доносиоца одлуке.

Циљ докторског рада је да се уведе нов начин моделовања сличности коришћењем интерполативне Булове алгебре и његова примена за решавања проблема и унапређење система за подршку одлучивању.

1.2. Полазне хипотезе

У докторском раду пошло се од следећих хипотеза.

Основне хипотеза:

- Логички приступ моделује сличност на начин који је близак људском расуђивању.
- Алати за подршку одлучивању могу се унапредити увођењем логичких мера сличности.

Помоћне хипотезе:

- Логичке релације импликације и еквиваленције могу се користити за мерење сличности између објеката.
- Релација еквиваленције заснована на интерполативној Буловој алгебри може се дефинисати као мера сличности.
- Мера сличности заснована на ИБА еквиваленцији може се уопштити увођењем параметра - параметарска ИБА еквиваленција.
- Сличност се може моделовати помоћу псеудо-логичке агрегације и мера сличности заснованих на ИБА.
- Логичке мере сличности засноване на ИБА могу се користити за моделовање консензуса.

- На основу ИБА логичких мера сличности могу се образовати различити класификатори.

1.3. Структура рада

У првом поглављу је идентификован проблем, предмет и циљ истраживања, и дефинисане су хипотезе.

У другом поглављу дата су основна теоријска полазишта која ће у даљем раду бити неопходна за разумевање сличности, као и увођење нових мера сличности. Прво су наведене основне дефиниције из класичне Булове логике и фази логике. Такође, наведене су и основне математичке дефиниције сличности и растојања. Затим су наведене постојеће логичке мере сличности засноване на фази би-импликацији. Посебна пажња је посвећена интерполативној Буловој алгебри која је значајна за дефинисање нове логичке мере сличности засноване на ИБА еквиваленцији.

У трећем поглављу се анализирају приступи мерењу сличности који су изучавани у литератури. Затим је дат критички осврт на метричке приступе, како би се утврдиле потребе и могућности за увођење нових. Посебна пажња је посвећена мерама сличности заснованим на вишевредносној логици као и параметризованим мерама сличности.

У четвртом поглављу дефинисана је нова логичка мера сличности заснована на ИБА еквиваленцији. Затим је објашњена основна идеја мерења сличности помоћу ИБА еквиваленције и дата њена графичка интерпретација. Примена ИБА еквиваленције је показана на илустративном примеру код закључивања на основу случајева. Поред тога поређење нове логичке мере сличности и познатих фази би-импликација приказано је на проблему моделовање консензуса у групном одлучивању.

У петом поглављу уопштена је ИБА еквиваленција увођењем параметара и приказано је њено основно тумачење код мерења сличности. Затим су проверена основна математичка својства мере сличности.

У шестом поглављу описују се познати алгоритми класификације најближи центроид и *k*-најближих суседа, који у даљем истраживању служе као основа за дефинисање логичког приступа класификацији.

У седмом поглављу дефинише се поступак за валидацију нових параметарских мера сличности помоћу логичке класификација засноване на ИБА. У другом делу врши се експеримент и анализирају резултати класификације на медицинским подацима.

У осмом поглављу примењује се логичка класификација у предвиђању банкротства средњих предузећа у Србији. Приказани су најбољи резултати класификације добијени за најближи центроид и *k*-најближих суседа заснованих на непараметарској или параметарској ИБА еквиваленцији и оператору логичке агрегације.

Последње поглавље је закључак, са посебним освртом на постављене хипотезе, преглед доприноса дисертације, као и на могуће правце даљег истраживања.

На крају рада је наведена литература коришћена приликом израде дисертације.

2. ОСНОВНЕ ТЕОРИЈСКЕ ПОСТАВКЕ

У другом поглављу су представљена основна теоријска полазишта из класичне и фази логике, интерполативне Булове алгебре и мерења сличности која ће у даљем раду бити неопходна за разумевање и увођење нових мера сличности.

2.1. Класична логика

Класична Булова логика је најпознатија стандардна логика. Највише се проучава у науци и представљала је полазну основу за развој многих других некласичних логика као што су вишевредносне логике, укључујући и фази логику. У овом поглављу наведене су основне дефиниције релације импликације и еквиваленције које су релевантне за тему истраживања.

Постоје два основна начина за дефинисање импликације у Буловој алгебри *(Baczynski & Jayaram, 2008a):*

Дефиниција 1. Нека је $\langle L, \vee \wedge \neg \rangle$ Булова алгебра. Релација $p \leftarrow q$ се назива *релацијом импликације* за свако $p, q \in L$ ако је

$$p \leftarrow q = \neg p \vee q .$$

Дефиниција 2. Нека је $\langle L, \vee \wedge \neg \rangle$ Булова алгебра. Релација $p \leftarrow q$ се назива *релацијом импликације* за свако $p, q \in L$ ако је

$$p \leftarrow q = \max \{ t \in L | p \wedge t \le q \},$$

где се $\le$ дефинише као $p \le q$ ако и само ако $p \vee q = q$ за свако $p, q \in L$.

Дефиниција 3 *(Bourbaki, 2004).* Нека је X скуп. *Релација еквиваленције* на скупу X је подскуп $R \subseteq X \times X$ који задовољава следећа три својства:

- Рефлексивност: за свако $x \in X$, $(x, x) \in R$
- Симетричност: за свако $x, y \in X$, ако $(x, y) \in R$, онда $(y, x) \in R$

- Транзитивност: за свако $x, y, z \in X$, ако $(x, y) \in R$ и $(y, z) \in R$, онда $(x, z) \in R.$

2.2. Фази логика

Фази логика је генерализација класичне логике јер омогућава да истинитосна вредност буде било која реална вредност из јединичног интервала. У традиционалној дуалној логици, исказ може бити тачан или нетачан, и ништа осим тога. Фази скупове увео је *Zadeh (1965)* са циљем да омогући математичко представљање и третирање неодређених и непрецизних информација. За разлику од традиционалних скупова где елемент или припада скупу или припада његовом комплементу, у теорији фази скупова елементи могу припадати више фази скупова истовремено. Теорија фази скупова је конструисана да служи као мост између природног језика и формалних модела као и за моделовање неизвесности која није стохастичке природе *(Zimmermann, 2010).*

Као што је познато, код класичних скупова карактеристична функција додељује вредности 1 или 0 сваком елементу универзалног скупа X . Ова функција се генерализује код фази скупова тако да функција припадности додељује сваком елементу универзалног скупа X реалну вредност која припада опсегу од 0 до 1.

Дефиниција 4 *(Klir & Yuan, 1995).* Нека је X универални скуп. *Фази скуп* A се дефинише функцијом припадности μ_A на следећи начин:

$$\mu_A : X \to [0,1]$$

где вредност ове функције указује на степен припадности елемента $x \in X$ посматраном фази скупу A.

У наставку поглавља се наводе дефиниције које служе као основа за конструисање многих логичких и параметарских мера сличности.

Дефиниција 5 *(Klement et al., 2000).* Функција $T:[0,1]^2 \rightarrow [0,1]$ се назива *троугаона норма* (*t-норма)* ако задовољава следеће услове за свако $x, y, z \in [0,1]$:

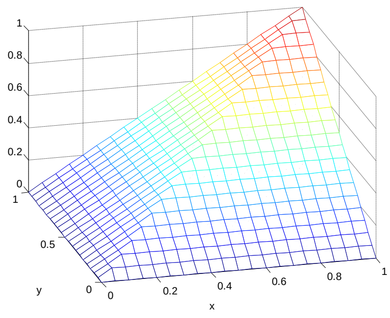
- Комутативност: $T(x,y) = T(y,x)$
- Асоцијативност: $T(x,T(y,z)) = T(T(x,y),z)$
- Монотоност: $y \leq z \Rightarrow T(x,y) \leq T(x,z)$
- Неутрални елемент 1: $T(x,1) = x \wedge T(1,y) = y$.

У фази логици *t*-норме су уопштење класичне двовредносне логичке конјункције. Класична Булова конјункција је такође комутативна и асоцијативна. Својство монотоности обезбеђује да се степен истинитости конјункције не смањује ако се истинитосне вредности у саставу конјункције повећавају. Захтев да јединица буде неутрални елемент одговара интерпретацији да је 1 тачна односно 0 нетачна вредност *(Klement et al., 2000).*

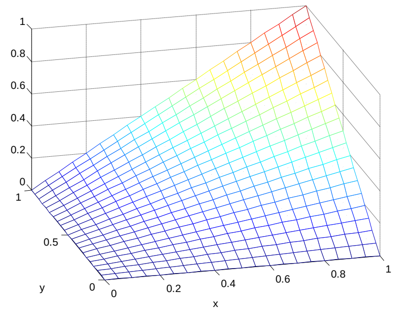
Постоји бесконачно много *t*-норми, а основне и најчешће коришћене *t*-норме су:

- Минимум: $T_M(x,y) = \min(x,y)$
- Производ: $T_P(x,y) = x \cdot y$
- Лукашијевич: $T_L(x,y) = \max(x+y-1,0)$
- Драстични производ: $T_{DP}(x,y) = \begin{cases} x, & y=1 \\ y, & x=1 \\ \textit{иначе} & 0 \end{cases}$.

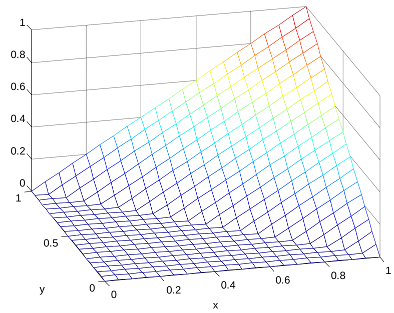
Минимум, алгебарски производ, Лукашијевич и драстични производ *t*-норма су приказани на Сликама 1-4.



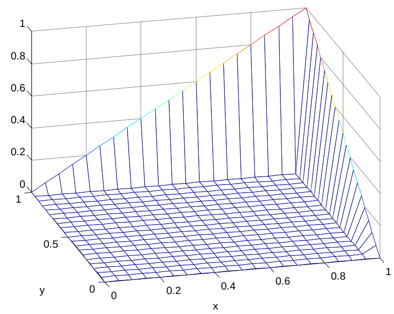
Слика 1. Минимум *t*-норма



Слика 2. Производ *t*-норма



Слика 3. Лукашијевич *t*-норма



Слика 4. Драстични производ *t*-норма

Поред основних троуганих норми у фази логици се проучавају и развијају бројне параметарске фамилије *t*-норми. Могуће је конструисати различите *t*-норме, било експлицитно путем дефиниције или трансформацијом претходно познатих функција. Тиме се обезбеђује класа *t*-норми са одређеним својствима која би се користила при решавању одговарајућих практичних проблема. Један од најчешћих начина конструисања *t*-норми је дефинисање параметарских класа *t*-

норми које садрже параметар из одређеног домена *(Klement et al., 2000)*. У оквиру овог поглавља наведене су неке од значајних параметарских *t*-норми у Табели 1.

Табела 1. Познате параметарске фамилије *t*-норми

Фамилија	***t*-норма**
Hamacher	$T_H(x,y)=\begin{cases} T_{DP}(x,y), & \lambda=\infty \\ 0, & \lambda=x=y \\ \dfrac{xy}{\lambda+(1-\lambda)(x+y-xy)}, & \lambda\in[0,\infty[\ \wedge(\lambda,x,y)\neq(0,0,0) \end{cases}$
Dombi	$T_D(x,y)=\begin{cases} T_{DP}(x,y), & \lambda=0 \\ T_M(x,y), & \lambda=\infty \\ \left(1+\left(\left(\dfrac{1-x}{x}\right)^{\lambda}+\left(\dfrac{1-y}{y}\right)^{\lambda}\right)^{\frac{1}{\lambda}}\right)^{-1}, & \lambda\in[0,\infty[\end{cases}$
Yager	$T_Y(x,y)=\begin{cases} T_{DP}(x,y), & \lambda=0 \\ T_M(x,y), & \lambda=\infty \\ \max\left(1-\left((1-x)^{\lambda}+(1-y)^{\lambda}\right)^{\frac{1}{\lambda}},\ 0\right), & \lambda\in[0,\infty[\end{cases}$
Frank	$T_F(x,y)=\begin{cases} T_M(x,y), & \lambda=0 \\ T_P(x,y), & \lambda=1 \\ T_L(x,y), & \lambda=\infty \\ \log_{\lambda}\left(1+\dfrac{(\lambda^{x}-1)(\lambda^{y}-1)}{\lambda-1}\right), & \lambda\in[0,1[\ \cup\ ]1,\infty] \end{cases}$
Schweizer-Sklar	$T_F(x,y)=\begin{cases} T_M(x,y), & \lambda=-\infty \\ T_P(x,y), & \lambda=0 \\ T_{DP}(x,y), & \lambda=\infty \\ \left(\max\left(x^{\lambda}+y^{\lambda}-1,0\right)\right)^{\frac{1}{\lambda}}, & \lambda\in[-\infty,0[\ \cup\ ]0,\infty[\end{cases}$

Проширена листа параметризованих фамилија *t*-норми може се пронаћи у (*Mizumoto, 1989; Lowen, 1996*) и код *(Klement et al., 2000)* где се детаљно проучавају њихове карактеристике. У примени и одабиру параметризованих фамилија значајну улогу има и брзина израчунавања (*Eckhardt et al., 2009).*

Фази импликација представља уопштење класичне импликације на цео јединични интервал $[0,1]$. Занимљиво је да су и поред различитих формула наведене две дефиниције (*Дефиниције 1* и *2*) еквивалентне у класичној Буловој алгебри. Са друге стране, у фази логици, где истинитосне вредности могу да узимају вредности из целог јединичног интервала, природне генерализације претходних дефиниција нису еквивалентне. Тиме се добијају две најпознатије класе фази импликација: *(S,N)*- и *R*- импликације *(Baczynski & Jayaram, 2008b).*

У овом раду усвојићемо следећу дефиницију за фази импликацију коју су међу првима дали *Fodor* & *Roubens (1994)*, а касније усвојили и *Baczynski & Jayaram (2008a).*

Дефиниција 6 *(Baczynski & Jayaram, 2008b).* Функција $I:[0,1]^2 \to [0,1]$ се назива *фази импликација* ако задовољава следеће аксиоме за свако $x, x_1, x_2, y, y_1, y_2 \in [0,1]$:

- Ако је $x_1 \le x_2$, онда $I(x_1, y) \ge I(x_2, y)$
- Ако је $y_1 \le y_2$, онда
- $I(x, y_1) \le I(x, y_2)$
- $I(0,0) = 1$
- $I(1,1) = 1$
- $I(1,0) = 0$.

Фази имликације засноване на претходно наведеним *t*-нормама су:

- Импликација заснована на минимум *t* -норми: $I_M(x,y) = \begin{cases} y, & x > y \\ 1, & \textit{иначе} \end{cases}$
- Импликација заснована на продизвод *t*-норми: $I_P(x,y) = \begin{cases} y/x, & x > y \\ 1, & \textit{иначе} \end{cases}$
- Импликација заснована на Лукашијевич *t*-норми: $I_L(x,y) = \min(x - y + 1, 1)$.

Коришћењем фази импликације може се генерисати релација фази би-импликације, која представља фази импликацију у оба смера.

Дефиниција 7 *(Callejas et al., 2012).* Функција $B:[0,1]^2 \to [0,1]$ се назива *фази би-импликација* ако задовољава следеће аксиоме за свако $x, y, z, w \in [0,1]$:

- B-комутативност: $B(x, y) = B(y, x)$
- B-идентитет: $B(x, x) = 1$
- $B(0,1) = 0$
- Ако је $w \le x \le y \le z$ онда $B(w, z) \le B(x, y)$.

Опште познате фази би-импликација су:

- Минимум би-импликација: $B_M(x, y) = \begin{cases} 1, & x = y \\ \min(x, y), & \textit{иначе} \end{cases}$
- Производ би-импликација: $B_P(x, y) = \begin{cases} 1, & x = y \\ \dfrac{\min(x, y)}{\max(x, y)}, & \textit{иначе} \end{cases}$
- Лукашијевич би-импликација: $B_L(x, y) = 1 - |x - y|$.

Фази би-импликација се интерпретира као $B(x, y) = T(I(x, y), I(y, x))$, где се подразумева да је $T(x, y)$ *t*-норма и $I(x, y)$ одговарајућа фази импликација.

2.3. Мера сличности и растојања

Математичке појмове метричких простора и метрика растојања увели су *Fréchet (1906)* и *Hausdorff (1914)* пре једног века. У овом поглављу наведене су основне математичке дефиниције сличности и растојања.

Дефиниција 8 *(Deza & Deza, 2014).* Нека је X скуп. Функција $s: X^2 \to R$ се назива *сличност* на скупу X ако за свако $x, y \in X$ важи:

- Ненегативност: $s(x, y) \ge 0$
- Симетричност: $s(x, y) = s(y, x)$

- Ограниченост са горње стране: $s(x,y) \le s(x,x)$ са једнакошћу ако и само ако је $x = y$.

Дефиниција 9 *(Deza & Deza, 2014).* Нека је X скуп. Функција $d: X^2 \to R$ се назива *растојање* на скупу X ако за свако $x, y \in X$ важи:

- Ненегативност: $d(x,y) \ge 0$
- Симетричност: $d(x,y) = d(y,x)$
- Рефлексивност: $d(x,x) = 0$.

Најчешће трансформације које се користе за добијање растојања d на основу сличности s, која је ограничена са горње стране са 1, су: $d = 1 - s$, $d = (1-s)/s$, $d = \sqrt{1-s}$, $d = \sqrt{2(1-s^2)}$, $d = \arccos s$, $d = -\ln s$, итд. *(Deza & Deza, 2014).* Посматрано са аспекта логике, растојање је негација сличности $d = \neg s$.

Дефиниција 10 *(Deza & Deza, 2014).* Нека је X скуп. Функција $d: X^2 \to R$ се назива *метрика* на скупу X ако је функција d растојање на X, и ако за свако $x, y, z \in X$ важи:

- Неједнакост троугла: $d(x,y) \le d(x,z) + d(z,y)$.

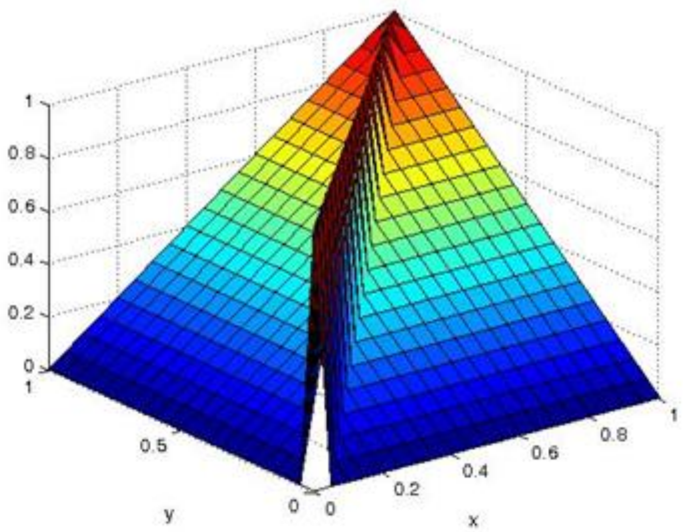
2.4. Мере сличности засноване на фази би-импликацији

У овом поглављу су наведене неке од познатих мера сличности заснованих на фази би-импликацијима. У радовима *(Milošević et al., 2013; Poledica et al., 2015)* показано је да минимум, производ и Лукашијевич би-импликације задовољавају својства мера сличности.

Сличност заснована на минимум би-импликацији је:

$$S_M(x,y) = \begin{cases} 1, & x = y \\ \min(x,y), & \textit{иначе} \end{cases} \tag{2.1}$$

чија је графичка интерпретација дата на Слици 5.



Слика 5. Мера сличности заснована на минимум би-импликацији

Теорема 1. Нека је $S_M:[0,1]^2 \to [0,1]$ минимум би-импликација. За било које $x, y \in [0,1]$ $S_M(x, y)$ је мера сличности.

Доказ: Потребно је показати да S_M задовољава својства дата у *Дефиницији 8*:

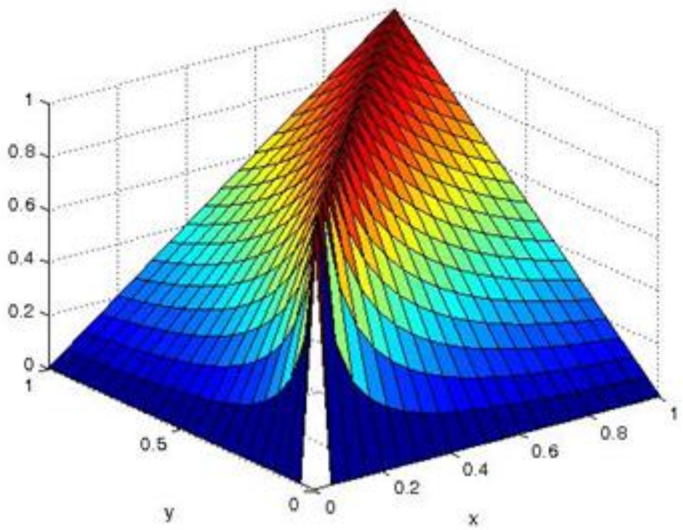
- Ненегативност: 1. $x = y = 0 \Rightarrow S_M(x,y) = 1 > 0$

  2. $x \neq y \Rightarrow \min(x,y) = S_M(x,y) \geq 0$
- Симетричност: $S_M(x,y) = \min(x,y) = \min(y,x) = S_M(y,x)$
- Ограниченост са горње стране: $S_M(x,x) = 1 > \min(x,y) = S_M(x,y)$ за $x \neq y$.

Сличност заснована на производ би-импликацији је:

$$S_P(x,y) = \begin{cases} 1, & x = y \\ \dfrac{\min(x,y)}{\max(x,y)}, & \textit{иначе} \end{cases} \tag{2.2}$$

и њена графичка интерпретација дата је на Слици 6.



Слика 6. Мера сличности заснована на производ би-импликацији

Теорема 2. Нека је $S_P:[0,1]^2 \to [0,1]$ производ би-импликација. За било које $x, y \in [0,1]$, $S_P(x, y)$ је мера сличности.

Доказ: Потребно је показати да S_P задовољава својства дата у *Дефиницији 8*:

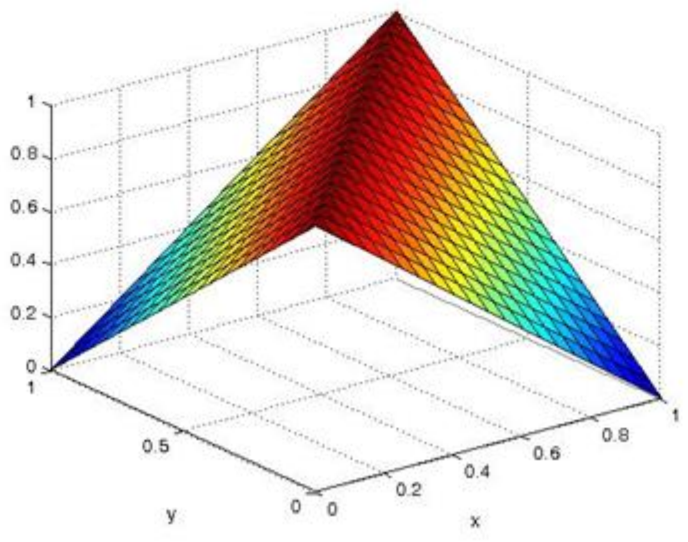
- Ненегативност: 1. $x = y = 0 \Rightarrow S_P(x,y) = 1 > 0$

  2. $\min(x,y) \le \max(x,y) \wedge x \ne y \Rightarrow \dfrac{\min(x,y)}{\max(x,y)} = S_P(x,y) \ge 0$
- Симетричност: $S_P(x,y) = \dfrac{\min(x,y)}{\max(x,y)} = \dfrac{\min(y,x)}{\max(y,x)} = S_P(y,x)$
- Ограниченост са горње стране: $S_P(x,x) = 1 > \dfrac{\min(x,y)}{\max(x,y)} = S_P(x,y)$ за $x \ne y$.

На крају се наводи сличност заснована на Лукашијевич би-импликацији:

$$S_L(x,y) = 1 - |x - y| \tag{2.3}$$

која је приказана на Слици 7.



Слика 7. Мера сличности заснована на Лукашијевич би-импликацији

Теорема 3. Нека је $S_L : [0,1]^2 \to [0,1]$ Лукашијевич би-импликација. За било које $x, y \in [0,1]$, $S_L(x, y)$ је мера сличности.

Доказ: Потребно је показати да S_L задовољава својства дата у *Дефиницији 8*:

- Ненегативност: $|x-y| \leq 1 \Rightarrow 1-|x-y| = S_L(x,y) \geq 0$
- Симетричност: $S_L(x,y) = 1-|x-y| = 1-|y-x| = S_L(y,x)$
- Ограниченост са горње стране: $S_L(x,x) = 1-|x-x| = 1 > 1-|x-y| = S_L(x,y)$ за $x \neq y$.

2.5. Интерполативна Булова алгебра

Теоријска основа интерполативне Булове алгебре (ИБА) која је овде приказана потребна је за дефинисање нове логичке мере сличности засноване на ИБА еквиваленцији.

Интерполативна Булова алгебра *(Radojević, 2000)* је реално вредносна реализација Булове алгебре на јединичном интервалу. За разлику од фази логике где није испуњен принцип искључења трећег, ИБА је конзистентан оквир у смислу да у њему важе сви аксиоми Булове алгебре (нпр. искључење трећег, контрадикција, итд.).

Увођењем интензитета релације односно градације Булових променљивих и функција, ИБА је погодна за многе реалне проблеме захваљујући дескриптивности градације (*Radojević, 2000*). У ИБА оквиру, логичке Булове функције се једнозначно преводе у одговарајуће генерализоване Булове полиноме (ГБП). Генерализовани Булови полиноми се састоје од елемената Булове алгебре који представљају променљиве, и од стандардних оператора сабирања +, одузимања -, и генерализованог производа $\otimes$.

Генерализовани производ је подкласа *t*-норми. Оператор генерализованог производа је било која функција која пресликава $\otimes:[0,1]^2 \to [0,1]$ и задовољава све четири аксиоме *t*-норми *(Дефиниција 4)* и додатни услов о ненегативности (*Radojević, 2008a)*:

$$\sum_{K \in P(\Omega/S)} (-1)^{|K|} \otimes a_i \geq 0 \tag{2.4}$$

Генерализовани производ се креће у следећем интервалу:

$$\max(a+b-1,0) \leq a \otimes b \leq \min(a,b) \tag{2.5}$$

Избор оператора за генерализовани производ не утиче на промену алгебре, односно алгебра је увек Булова. Са друге стране, у зависности од природе променљивих у ГБП-у, постоје карактеристични случајеви за избор одговарајућег оператора *(Radojević, 2008a; Radojević, 2008b; Dragović et al., 2014)*. На пример, функција минимума је одговарајућа у случају променљивих које имају исту или сличну природу, док се за независне променљиве користи оператор производ.

Поступак превођења Булових функција у ГБП је дефинисан код *Radojević (2008b)*: За примарне променљиве $\{a_1,\ldots a_n\}$

$$
\begin{aligned}
&\left(a_i \wedge a_j\right)^{\otimes} = \begin{cases} a_i \otimes a_j, & i \neq j \\ a_i, & i = j \end{cases}, \\
&\left(a_i \vee a_j\right)^{\otimes} = a_i + a_j - \left(a_i \wedge a_j\right)^{\otimes}, \\
&\left(\neg a_i\right)^{\otimes} = 1 - a_i
\end{aligned} \tag{2.6}
$$

За комбиноване елементе $F\left(a_1,\ldots,a_n\right), G\left(a_1,\ldots,a_n\right) \in BA(\mathbf{\Omega})$

$$
\begin{aligned}
&\left(F \wedge G\right)^{\otimes} = F^{\otimes} \otimes G^{\otimes}, \\
&\left(F \vee G\right)^{\otimes} = F^{\otimes} + G^{\otimes} - \left(F \wedge G\right)^{\otimes}, \\
&\left(\neg F\right)^{\otimes} = 1 - F^{\otimes}.
\end{aligned} \tag{2.7}
$$

Софтверска подршка за поступак превођења Булових функција у ГБП је представљена у *Milosević et al. (2014)*.

Овде је приказано превођења Булове функције за еквиваленцију $a \Leftrightarrow b = \left(a \Rightarrow b\right) \wedge \left(b \Rightarrow a\right)$ у одговарајући генерализован Булов полином. У првом кораку се врши трансформација на *структурном* нивоу:

$$
\begin{aligned}
\left(a \Leftrightarrow b\right)^{\otimes} &= \left(\left(a \Rightarrow b\right) \wedge \left(b \Rightarrow a\right)\right)^{\otimes} = \\
&= \left(a \Rightarrow b\right)^{\otimes} \otimes \left(b \Rightarrow a\right)^{\otimes} = \\
&= \left(1 - a + a \otimes b\right) \otimes \left(1 - b + a \otimes b\right) = \\
&= 1 - b + a \otimes b - a + a \otimes b - a \otimes a \otimes b + a \otimes b - a \otimes b \otimes b + a \otimes b \otimes a \otimes b = \\
&= 1 - b + a \otimes b - a + a \otimes b - a \otimes b + a \otimes b - a \otimes b + a \otimes b = \\
&= 1 - b - a + 2 \cdot a \otimes b
\end{aligned} \tag{2.8}
$$

Након структурне трансформације добијени израз се преноси на *вредносни* ниво. Добијени генерализовани Булов полином у ИБА-а једнозначно одговара полазној релацији еквиваленције. Такође добијени ИБА полином задовољава својства рефлексивности, симетричности и транзитивности само у случају када се функција минимума изабере за генерализован производ, односно $\otimes := \min$ *(Radojević, 2010)*. Тиме се за ИБА еквиваленцију добија следећи израз на *вредносном* нивоу:

$$
\left(a \Leftrightarrow b\right)^{\otimes} = 1 - a - b + 2 \cdot \min\left(a, b\right) \tag{2.9}
$$

Како је минимум оператор одговарајући у случају променљивих које имају исту или сличну природу, следи да је конзистентно поређење различитих објеката једино могуће ако се поређење врши по истом критеријуму *(Radojević, 2010)*.

2.5.1. Логичка агрегација заснована на ИБА

Логичка агрегација (ЛА) је конзистентан и транспарентан поступак заснован на ИБА који се користи за агрегацију атрибута *(Radojević, 2008a)*. Основна идеја је да се омогући дескриптивност увођењем логичких функција као оператора агрегације. Логичке функције се представљају у форми логичких израза који се преводе у одговарајуће генерализоване Булове полиноме. ЛА метода има два корака *(Radojević, 2008a)*:

- Први корак је нормализација атрибута $\|\bullet\|:\Omega \to [0,1]$
- Други корак подразумева агрегацију нормализованих вредности у једну резултујућу вредност коришћењем оператора логичке агрегације $Aggr[0,1]^n \to [0,1]$.

ЛА може бити у форми псеудо-логичке функције која представља линеарно конвексну комбинацију генерализованих Булових полинома. Може се изразити у облику других познатих оператора агрегације као што су аритметичка средина, тежинска сума, минимум или максимум оператори, *OWA* оператор, дискретан *Choquet* интеграл, генерализован дискретан *Choquet* интеграл што је детаљно представљено у раду *(Radojević, 2008a)*. Осим тога било која задата логичка функција од стране експерта може се користити за агрегацију атрибута. Због тога је ЛА дескриптиван и флексибилан алат којим је могуће конструисати различите функције агрегације у зависности од проблема (*Milošević et al., 2013; Poledica et al., 2010)*. Поред тога омогућава и моделовање интеракције између фактора који се агрегирају *(Radojević & Petrović, 2001; Radojević, 2008a)*.

Предности ЛА су представљали основу за конструисање класификатора са експертским логичким функцијама у поглављу 8.

3. КРИТИЧКИ ОСВРТ НА МЕРЕЊЕ СЛИЧНОСТИ

У трећем поглављу детаљно ће се анализирати приступи мерењу сличности који су изучавани у литератури. Затим, биће дат критички осврт на релевантне приступе како би се утврдиле могућности и потребе за увођењем нових приступа. Посебна пажња биће усмерена на мере сличности засноване на логичким изразима као и на параметарске мере сличности.

3.1. Сличност и различитост

Сличност, односно различитост, представља један од фундаменталних концепата који се налази у основи целокупног људског расуђивања и искуства. У свакодневном животу сличност обично подразумева одређени степен близине два физичка објекта или идеје, нпр. дужина, временски интервал, разлика у рангу, итд., док се појам метрика често користи као стандард за мерење *(Petrović, 1998; Deza & Deza, 2014)*. Мерење сличности/различитости је важно за решавање великог броја проблема и представља основни захтев у скоро свим применама које укључују класификацију, кластеровање, селекцију карактеристика, детекцију изузетака, регресију и претрагу. Сличност и различитост између објеката се могу мерити само на основу неколико карактеристика. Често је потребно извршити селекцију карактеристика пре самог израчунавања сличности између објеката. Ако можемо измерити сличност/различитост онда (*Bandyopadhyay & Saha, 2013*; *Teknomo, 2015*):

- Можемо разликовати један објекат у односу на други;
- Можемо груписати објекте на основу сличности/различитости;
- Када их групушемо, можемо разумети карактеристике сваке групе;
- Можемо објаснити понашање група објеката;
- Груписање такође може дати ефикаснију организацију и повраћај информација;
- Можемо класификовати нови објекат у неку групу;
- Можемо предвидети понашање новог објекта;

- Можемо поједноставити податке које поседујемо успостављањем смислених веза у области анализе података;
- Можемо открити структуру у оквиру неког скупа података;
- Можемо предузети/спровести акцију, план и одлуку на основу структуре и предвиђања података.

Број постојећих метрика сличности је бесконачан и не постоји јасно дефинисан поступак како одабрати „добру“ односно одговарајућу функцију сличности за одређени проблем.

3.2. Постојећи правци

Бројне дефиниције мера сличности су предложене у литератури. Неке од њих су ограничене на посебан домен, док су друге широко распрострањене и примењене готово у свакој дисциплини. Важно је напоменути да не постоји строга математичка дефиниција мере сличности. Најчешће постоји само опште разумевање њене интерпретације - што је већа вредност, то су сличнији објекти. Због свега тога у домену примене веома је сложено питање избора једне или више мера које су најприкладније за одређени задатак *(Deza & Deza, 2014)*.

Према (*Bandyopadhyay & Saha, 2012*) мерење сличности зависи од типа података и може се поделити у неколико категорија:

- Мере сличности за бинарне променљиве;
- Мере сличности за категоријске променљиве;
- Мере сличности за ординалне променљиве;
- Мере сличности за квантитативне променљиве;
- Различитост између две групе променљивих.

Успешност било које технике анализе података зависи од избора мере сличности. Сам избор функције сличности зависи од природе података који су доступни за решавање проблема. Према (*Deza & Deza, 2014*), постоје функције сличности за нумеричке податке, за бинарне податке, сличности засноване на Еуклидском растојању и сличности засноване на корелацији.

Сличност/различитост се најчешће мери функцијама растојања у метричким просторима. Критика метричких својстава при креирању функција сличности за различите примене представља полазну мотивацију за истраживање у овом раду. Овде су посебно важни приступи мерењу сличности где су улазни подаци фази бројеви или било које друге вредности (квантитативне или квалитативне) које се могу свести на јединични интервал. Постоји више начина за поређење фази вредности.

Blosh (1999) прави разлику између дистанци које пореде само функције припадности које представљају фази објекте и дистанци које комбинују и просторну удаљеност између објеката и између функција припадности којих има мање у литератури. У истом раду аутор систематизује главне приступе при дефинисању фази дистанци. Први приступ се ослања на дефиницију *crisp* дистанце и покушава да је генерализује *(Dubois & Jaulent, 1987; Bloch & Maitre, 1995a; Bloch & Maitre, 1995b)*. Друга група закључује о дистанци из функције сличности *(Zadeh, 1971; Wang et al., 1995; De Baets, & Mesiar, 1997),* док трећи приступ изводи дистанцу на основу односа између два скупа *(Bouchon-Meunier et al., 1996)*. Последња група су симболички приступи који дефинишу лингвистичке променљиве које представљају дистанце *(Krishnapuram et al., 1993).*

Сличне групе поређења фази вредности даје и *Le Capitaine (2012).* Код њега се први приступ заснива на широкој класи мера једнакости у чијој је основи мера растојања функција припадности код фази концепата. Другу групу чине мере засноване на скуповним операцијама *(Tversky, 1977; Bouchon-Meunier et al., 1996)* које су проширене за рад са фази скуповима (фази пресек, унија, кардиналност). Трећи начин за дефинисање мера сличности се односи на коришћење логичких концепата односно фази импликације, као што је предложено код *(Bandler & Kohout, 1980; Hirota & Pedrycz, 1991).*

У литератури је предложено неколико метода за добијање мера сличности које се заснивају на функцији растојања. Када се растојање повећава, сличност би требало да се смањује. Различите функција зависности сличности и растојања су предложене *(e.g. Zimmermann & Zysno, 1980; Shepard, 1987)*. Међутим, добијање мере сличности помоћу мере растојања имплицитно подразумева да сличност задовољава својства метрике, што није јасно видљиво у практичној примени. Као последицу тога истраживачи у пракси више воле да користе неметричке моделе сличности *Le Capitaine (2012)*. Критика метричких приступа биће дата касније у раду.

Други начин за поређење фази вредности потиче од основних разматрања код теорије скупова где се дефинишу унија, пресек и комплемент за фази скупове. Док се мере засноване на метрици могу интерпретирати као близина (*енг.* proximity) фази скупова, мере засноване на теорији скупови се посматрају као приближна једнакост *(Le Capitaine, 2012)*. Детаљан опис мера поређења и сличности које су засноване на фази теорији скупова дат је код *Cross & Sudkamp (2002)*.

Zwick et. al (1987) су направили компаративну анализу мера сличности фази концепата које се заснивају на геометријској дистанци или на теорији скупова. Анализа је вршена на проблему лингвистичке апроксимације. За грубу категоризацију фази концепата све мере су се добро показале, али су неке биле супериорније да одреде степен сличности. Занимљиво је да су се најбоље показале мере које хватају само део функције припадности. Као најбољи избор на дужи рок и за шири спектар примена показале су се неке од прилагођених верзија *Hausdorff* метрике и индекса неконзистенстости. Детаљније резултате осталих разматраних мера могу се пронаћи у *(Zwick et. al, 1987)*.

Последњи разматрани приступ се заснива на посматрању степена импликације елемената који припадају A у односу на елементе из B и обратно *(Bandler & Kohout, 1980)*. Мере сличности засноване на логичком поређењу могу се пронаћи код *(Saastamoinen, 2006; Le Capitaine & Frelicot, 2009b)*.

У наставку рада после критичког осврта постојећих праваца дат је детаљан преглед мера сличности заснованих на вишевредносним импликацијама и еквиваленцијама.

3.3. Критички осврт

Критички осврт на постојеће приступе мерењу сличности се највећим делом односи на метричке приступе. Тврђење да мере сличности морају да задовољавају својства метричких простора преиспитивано је у многим радовима *(Tversky, 1970; Tversky, 1977; Santini & Jain, 1999; Cock & Kerre, 2003; Klawonn, 2003; Le Capitaine & Frelicot, 2009a).* Као резултат у домену примене јављају се бројни модели сличности који не захтевају испуњење метричких постулата нпр. *(Saastamoinen, 2006).*

У основним дефиницијама мера и метрика сличности (*Дефиниције 8, 9 и 10*) већина наведених својстава се могу подвргнути дискусији. Због тога је потребно нагласити да не постоји строга дефиниција неке опште мере сличности која је одговарајућа за више различитих намена.

У научној литератури највише су присутни геометријски модели за мерење сличности. Код њих су објекти представљени као тачке у неком координатном простору тако да посматрана различитост између објеката одговара метричком растојању између одговарајућих тачака. Приступ заснован на метрикама има своје корене у радовима где је питање било како мерити удаљеност између две реалне функције. Код фази сличности, основна идеја је да се фази скуп посматра као тачка у векторском простору. Најчешће се користи општа форма Минковски p метрике дефинисане за $p \geq 1$:

$$d_p(A,B) = \left(\sum_{x \in X} \left| f_A(x) - f_B(x) \right|^p \right)^{1/p} \tag{3.1}$$

Из наведене метрике се изводе добро познате функције растојања: Хеминг (Менхетн) за $p=1$, Еуклидска за $p=2$ и Чебишевљева $p=\infty$. Може се запазити да за $p<1$, d_p није метрика јер не задовољава услов за неједнакост троугла. Хеминг и Еуклидска функција растојања се такође означавају као L_1 и L_2 норме. Повећавањем вредности p више се тежине ставља на разлику између вредности атрибута.

Сматра се да су све анализе растојања метричке по природи, иако неке (нпр. хијерархијско кластеровање) пре имају структуру стабла него димензионо организованих простора. Већина теоријских и емпиријских анализа сличности претпостављају да се објекти могу адекватно представити као тачке у неком координатном простору и да се различитости понашају као метричке функције растојања. Према *Tversky (1977)* обе претпоставке, димензионалност и метричка природа, су отворене за дискусију.

Да би се оценила адекватност геометријског приступа може се испитати валидност метричких аксиома мере сличности/различитости које су наведене у *Дефиницији 8, 9* и *10*.

Аксиома минималности имплицира да сличност између неког објекта и себе самог је иста за све објекте. Ова претпоставка не важи за неке мере сличности. На пример, вероватноћа да се два идентична објекта оцене као иста а не различита, није константна за све објекта. Односно, неки објекат се чешће идентификује као неки други објекат него као он сам. Ако се вероватноћа идентификације интерпретира као мера сличности онда ове опсервације нарушавају минималност и самим тим нису компатибилни са моделом растојања *(Tversky, 1977)*.

Сличност се посматра као основни пример релације симетричности. Заиста, претпоставка о симетричности се суштински наглашава код свих приступа сличности. Насупрот овој традицији, *Tversky (1977)* даје емпиријске доказе о постојању асиметричних сличности, и заступа тезу да сличност не треба третирати као релација симетричности. Кажемо да портрет личи на особу, а не да особа личи на портрет; кажемо да је елипса као круг, а не да је круг као елипса; кажемо да је Северна Кореја као Кина, а не да је Кина као Северна Кореја. Некада оба смера у релацији могу носити посебно значење. „Живот је као представа“ где се указује на то да људи играју улоге или „представа је као живот“ где се указује на то да нека представа може да ухвати суштинске елементе људског живота *(Tversky, 1977).*

Својство рефлексивности је такође тема за дискусију, јер постоје случајеви где сличност неке тачке са самом собом није 1 (*Le Capitaine & Frelicot, 2009a*). Међутим, својство транзитивности је највише подвргнуто критици.

Неједнакост троугла имплицира да ако је *a* веома сличан *b*, и *b* веома сличан *c*, онда *a* и *c* не могу бити веома различити један од другог. Можемо посматрати сличност између земаља: Јамајка је слична Куби (услед географске близине); Куба је слична Русији (због политичких тенденција); али Јамајка и Русија немају никакву сличност *(Tversky, 1977).* Овај пример показује да сличност, иако очекивано, није транзитивна. Међутим, овакви примери не морају неопходно да оповргну нејаднакост троугла, већ само да укажу на то да ова претпоставка не треба да се прихвата као камен темељац модела сличности *(Tversky, 1977).* Као закључак *Tversky (1977)* каже: „Иако постоји доста успешних примена, геометријски приступ у анализи сличности наилази и на неке потешкоће. Претпоставка о димензионалности има ограничену примењивост, метричке аксиоме се доводе у питање.“

Као алтернативу *Tversky (1977)* развија теоријски правац који се заснива на поређењу карактеристика. Овај правац није ни димензиони ни метрички по природи, а заснива се на теорији скупова. Овде су објекти представљени као колекције карактеристика, а сличност се описује као процес упаривање особина:

$$s(A,B) = \frac{A \cap B}{A \cap B + \alpha |A - B| + \beta |B - A|} \tag{3.2}$$

Сличност се повећава са повећавањем заједничких/општих карактеристика и/или смањењем карактеристика које нису заједничке (односно карактеристике које припадају једног објекту, а не другом). *Tversky (1977)* уводи и објашњава неколико емпиријских појмова: опште и разликујуће карактеристике, везе између оцењене сличности и растојања, асиметричне сличности као и утицај контекста на сличност.

Као генерализацију концепта еквиваленције *Zadeh (1971)* је увео релацију фази еквиваленције. Од тада се фази еквиваленција широко проучава као начин за мерење степен сличности или немогућности разликовања објеката у неком универзуму *(Ćirić et al., 2007)*. У зависности од проблема односно контекста у којима су се примењивале, фази еквиваленције су добијале разна имена као што су релације сличности *(Zadeh, 1971)*, оператори неразликовања (*енг.* indistinguishability operators) *(Valverde, 1985; Jacas, 1990; Jacas & Recasens, 1995; Demirci, 2003a)*, *t*-еквиваленције *(De Baets & Mesiar, 1998; De Baets & Mesiar, 2002)*, вишевредносне *M*-еквиваленције *(Demirci, 2003b; Demirci, 2003c)*, итд. У случају *t*-еквиваленција услов неједнакости троугла се замењује условом за *t*-транзитивност.

При моделовању сличности поставља се питање да ли је потребно да фази једнакост и сличност задовољавају транзитивност. Ову дискусију отворили су *De Cock & Kerre (2003)* који тврде да, у општем случају, фази релација еквиваленције није одговарајућа за моделовање приближне једнакости (*енг.* approximate equality), а која је фази еквивалент класичној једнакости, јер сматрају да је појам транзитивности у супротности са интуицијом. Они наводе пример тзв. Поенкаровог парадокса (*Poincaré, 1902; Poincaré, 1904*): дечак може бити сличан своме оцу, отац може бити сличан своме оцу, дечаковом деди, али и даље постоји могућност да дечак и његов деда немају ништа заједничко. Иако је овде интуитивно јасно да приближна једнакост није транзитивна, у теорији фази скупова приближна једнакост се обично моделује са фази релацијом еквиваленције која испуњава неку врсту транзитивности.

У свом раду *Klawonn (2003)* се детаљно бавио претходним проблемом али није дошао до коначног одговора. Иако није успео да одговори на ово питање, указао је на неке битне проблеме који се могу јавити при моделовању сличности. Тако он у свом раду наводи *Poincaré-a (1904)* који је приметио да немогућност разликовања често није транзитивно у стварном животу. Следећи пример који је дао *Klawonn (2003)* добро објашњава овај проблем. Претпоставимо да особа жели да купи луксузан аутомобил и види рекламу са специјалном понудом за један аутомобил од 29.995 долара. Одлучује се да купи аутомобил и када стигне код продавца открива да је цена уствари 29.996 долара. Наравно он неће променити своју одлуку само зато што је цена већа за 1 долар него што је очекивао. Ако итеративно уопштимо аргумент да можемо да повећавамо цену аутомобила за 1 долар без утицаја на купца, онда долазимо до закључка да купцу можемо наплатити било коју цену. Можемо рећи да се две цене не могу разликовати уколико је њихова разлика мања од једног долара. У овом примеру је показано како људи третирају појам величине у стварном животу. Ми често не посматрамо конкретне бројева него распоне бројева. Другим речима, прави се разлика између опсега који се често поистовећују са неким бројем.

Ако уопштимо претходни пример, у многим ситуацијама људи обично сматрају два броја немогућим за разликовање када њихова разлика не пређе одређену границу толеранцију ε. Следећа релација је рефлексивна и симетрична али није транзитивна:

$$x \approx y \Leftrightarrow |x - y| \leq \varepsilon \tag{3.3}$$

Критика метричких мера сличности отворила је простор за нове правце истраживања и моделовања сличности.

3.4. Логичке мере сличности

Мере сличности засноване на фази би-импликацији имају јаку математичку основу описану код *(Klawonn & Castro 1995; Formato et al. 1999)* и примењене су у многим областима. Фази би-импликација може бити заснована на различитим *t*-нормама (нпр. минимум, производ) и представља проширење класичне би-импликације тако да обухвата цео јединичан интервал $[0,1]$ *(Moser, 2006; Callejas et al., 2012).* Би-импликацију у чијој основи се налази минимум *t*-норма проучавали су *Di Nola et al. (1988)* док су се би-импликацијом заснованом на Лукашијевич *t*-норми бавили *Li & Jin (2012). Wang (1997)* је дефинисао веома дескриптивну меру сличности користећи производ и Лукашијевич би-импликацију и аритметичке средине за агрегацију. Код *(Luukka & Leppalampi, 2006; Luukka, 2011)* генерализована Лукашијевич структура је коришћена као функција сличности за проблеме класификације у медицини чиме су постигнути одлични резултати. Мера сличности заснована на Лукашијевич би-импликацији је такође примењена код система за контролу саобраћајних сигнала *(Niittymakia & Turunen, 2003).* Код *(Le Capitaine & Frelicot, 2009b)* се такође користи фази би-импликација. *De Baets & Mesiar (2002)* су показали да у случају *R*-импликација, би-импликација испуњава услове за *t*-једнакост ако и само ако је *t*-норма континуална са леве стране. Због тога, у свом раду *Le Capitaine (2012)* користи би-резидуалне односно *R*-импликације јер остале *S*-, *QL*- и *D*- импликације не испуњавају услов за *t*-једнакост. *Le Capitaine (2012)* је дефинисао оквир за фази мере сличности засноване на би-импликацијама и применио их успешно на реалним подацима. Код *Beliakov et al. (2014)* мере консензуса су конструисане коришћењем Лукашијевич би-импликације и *Bonferroni* агрегационе функције.

Saastamoinen (2006) се бавио истраживањем на тему мера поређења заснованих на вишевредносним алгебарским структурама. Конструисао је неколико логичких мера поређења заснованих на *t*-нормама и *t*-конормама *(Saastamoinen, 2004a),* параметризованим *S*-импликацијама и еквиваленцијама *(Saastamoinen, 2004b)*, и генерализованим *3П*-унинормама *(Saastamoinen & Sampo, 2004)* где је генерализована средња вредност коришћена као оператор агрегације.

3.5. Параметризоване мере сличности

У овом поглављу дат је преглед параметарских мера сличности које су значајне са аспекта овог истраживања. Увођење параметара у модел је начин да се пронађе мера сличности посебно прилагођена неком проблему односно подацима.

Tversky (1977) је још седамдесетих година прошлог века увео две параметарске мере за сличност између два скупа. Овде је основна идеја да је удаљеност већа уколико два фази скупа имају мањи пресек. Као специјалан случај, за вредности параметара $\alpha = \beta = 1$ добија се *Jaccard* индекс односно коефицијент сличности *(Cross, & Cabello, 1995)*. Постоје и многе друге параметризоване функције сличности које су инспирисане моделом *Tversky*-ог *(e.g. Cross, 2006; Jimenez et al. 2012).*

Другу групу чине параметарске мере сличности које се заснивају на параметарским *t*-нормама. У научној литератури параметри се најчешће уводе за решавање проблема класификације односно када се креирају различити класификатори засновани на сличности. Да би побољшали перформансе класификације *Le Capitaine & Frelicot (2009a)* разматрали су проблем одбацивања неких патерна услед тога што нису довољно поуздани јер припадају истовремено већем броју класа. У ту сврху *Le Capitaine & Frelicot (2009a)* су предложили конструисање параметарских мера неодређености заснованих на *R*-импликацијама и параметарским *Hamacher* и *Dombi t*-нормама чија вредност указује на степен у коме би патерн требало да се додели некој класи. Такође, *Le Capitaine (2012)* је предложио општи оквир за фази мере сличности засноване на *T*-једнакостима изведеним из *R*-импликација. У истом раду увео је модел који омогућава учење параметарских *R*-импликација као што су *Hamacher*, *Dombi*, *Frank*, *Yager*, *Schweizer-Sklar*, итд.

Трећи приступ представља конструисање различитих мера поређења где се параметри уводе не само код мере сличности већ и код оператора агрегације. *Saastamoinen (2004b)* је конструисао више мера поређења које користе параметарску *Kleene-Dienes*, *Reichenbach*, Лукашиевич, као и комбиновану Лукашиевич и *Shweizer-Sklar* импликацију, као и мере поређења са логичком еквиваленцијом заснованом на претходним импликацијама. Наведене импликације су параметризоване стављањем променљивих у експоненцијалну форму. За генерисање мера поређење додатно је увео тежине и генерализовану средњу вредност као оператор агрегације. Такође, у раду *(Saastamoinen, 2004a)* дефинисане су комбиноване мере поређења засноване на *t*-нормама и *t*-конормама, где се фази унија и пресек комбинују са генерализованим просеком и тежинама. Коришћене су параметризоване фамилије *t*-норми и *t*-конорми као што су *Dombi*, *Frank*, *Schweizer-Sklar*, *Yager* и *Yu*. *Frank* и *Schweizer-Sklar* *t*-норме су и копуле па имају добре статистичке карактеристике, док су *Yager*, *Dombi* и *Yu* *t*-норме међу најпопуларнијим фамилијама за моделовање пресека *(Saastamoinen, 2006)*. Након опсежног истраживања, *Saastamoinen (2006)* закључује да мере поређења засноване на еквиваленцији у већини случајева дају значајно боље резултате него мере засноване на импликацији. Такође, у свим резултатима класификације комбиноване мере поређења засноване на *Frank* *t*-норми показале су се приближно исте или боље од осталих.

Saastamoinen et al. (2002a; 2002b) су проучавали класификаторе који су засновани на обичној и генерализованој Лукашијевич структури као фази мери сличности. Оцена параметара је извршена у три фазе: прво су оптимизоване само тежине у обичној Лукашијевич структури, затим само коефицијент степена у оквиру генерализоване Лукашијевич структуре, а на крају и тежине и степен истовремено. На три различита скупа података показано је да класификација заснована на фази сличности даје боље резултате од традиционалног класификатора најближих суседа. При класификацији, за добар избор степена, генерализована Лукашијевич структура даје једнако добре или боље резултате од обичне. Увођење тежина такође унапређује класификацију, посебно када се на прави начин комбинује са генерализованом Лукашијевич структуром.

4. ЛОГИЧКА МЕРА СЛИЧНОСТИ - ИБА ЕКВИВАЛЕНЦИЈА

У овом поглављу уводи се нова логичка мера сличности заснована на ИБА еквиваленцији, објашњава се основна мотивација за њено увођење и проверају својства мере сличности. На примеру консензуса и закључивања на основу случајева се приказује апликативна вредност логичке мере сличности засноване на ИБА еквиваленцији.

4.1. Логичка мера сличности заснована на ИБА еквиваленцији

Када говоримо о мерењу сличности у ИБА оквиру, идеја је да се за мерење сличности између објеката користи релација ИБА еквиваленције. Релација еквиваленције се у општем случају дефинише у облику следећег логичког израза:

$$a_1 \Leftrightarrow a_2 = \left(a_1 \Rightarrow a_2\right) \wedge \left(a_2 \Rightarrow a_1\right) \tag{4.1}$$

Овај израз може да третира само две вредности: 0 и 1. Међутим, људска перцепција сличности природно у себи садржи градацију. Да би се одредио степен сличности, у литератури постоји низ различитих меких мера сличности. У овом раду уводи се ИБА еквиваленција као мера сличности која користе цео јединични интервал. Тако, ГБП који једнозначно одговара горе наведеном логичком изразу за еквиваленцију има следећи облик:

$$\left(a_1 \Leftrightarrow a_2\right)^{\otimes} = 1 - a_1 - a_2 + 2 \cdot a_1 \otimes a_2 \tag{4.2}$$

Будући да објекти могу поредити само по истој особини, генерализовани производ мора бити реализован као минимум:

$$S_{IBA}\left(a_1, a_2\right) = 1 - a_1 - a_2 + 2 \cdot \min\left(a_1, a_2\right) \tag{4.3}$$

За наведени израз проверавају се сва потребна својства која према *Дефиницији 8* треба да задовољи мера сличности. ИБА еквиваленције за мерење сличности прво је примењена код *(Poledica et al., 2009).* Провера њених метричких својстава започета је код *(Rakićević, 2012; Poledica et al., 2013; Milošević et al., 2013)*, а касније заокружена у раду *(Poledica et al. 2015)*.

Теорема 4 *(Poledica et al., 2015)*. Нека је $S_{IBA}:[0,1]^2 \to [0,1]$ релација еквиваленције у ИБА. За било које $x, y \in [0,1]$, $S_{IBA}(x, y)$ је мера сличности.

Доказ: Потребно је показати да S_{IBA} задовољава својства дата у *Дефиницији 8*:

Ненегативност: $S_{IBA}(x,y) \geq 0$

$S_{IBA}(x,y) = 1 - x - y + 2 \cdot \min(x,y) = 1 - (x + y - 2 \cdot \min(x,y))$

За $y > x$ израз постаје $x + y - 2 \cdot \min(x,y) = y - x \leq 1$, у супротном, $x + y - 2 \cdot \min(x,y) = x - y \leq 1$. Тиме је показано да је $S_{IBA}(x,y) = 1 - (x + y - 2 \cdot \min(x,y)) \geq 0$.

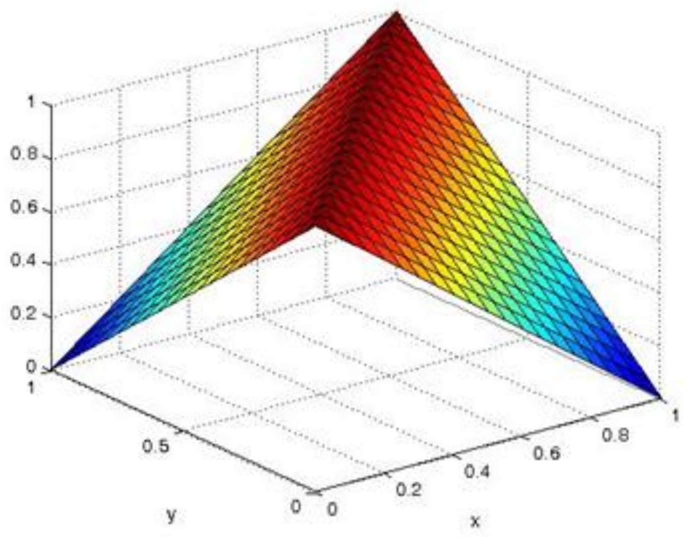
Симетричност: $S_{IBA}(x,y) = S_{IBA}(y,x)$

$S_{IBA}(x,y) = 1 - x - y + 2 \cdot \min(x,y) = 1 - x - y + 2 \cdot \min(y,x) = S_{IBA}(y,x)$, што важи услед комутативности функције минимума.

Ограниченост са горње стране: $S_{IBA}(x,y) \leq S_{IBA}(x,x)$ са једнакошћу ако и само ако је $x = y$

$S_{IBA}(x,x) = 1 - x - x + 2 \cdot \min(x,x) = 1 - 2 \cdot x + 2 \cdot x = 1 > S_{IBA}(x,y)$ за $x \neq y$, услед идемпотентности функције минимума за две исте вредности.

Мера сличности заснована на ИБА еквиваленцији приказана је на Слици 8.



Слика 8. Мера сличности заснована на ИБА еквиваленцији

4.2. Интерпретација ИБА мере сличности

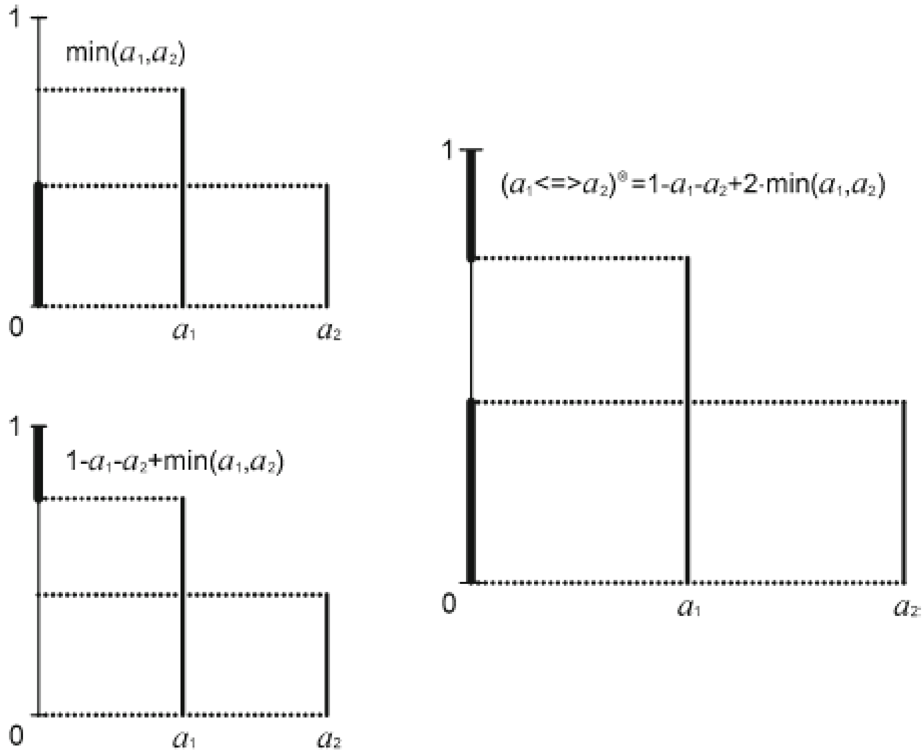
Основна мотивација за избор претходно наведене логичке функције за меру сличности дата је у овом одељку.

Релација еквиваленције изражена у познатој форми двоструке односно би-импликације дата у претходној једначини, може се написати и у другачијем облику који је логички идентичан претходном:

$$a_1 \Leftrightarrow a_2 = (a_1 \wedge a_2) \vee (\neg a_1 \wedge \neg a_2) \tag{4.4}$$

Приказан у овој форми наведени израз може да објасни само значење перцепције сличности које се заступа у овом раду, а истовремено представља основну мотивацију за разматрање ИБА еквиваленције за мерење сличности.

Према Радојевићу, графичка интерпретација ИБА еквиваленције као мере сличности дата је на Слици 9 *(Poledica et al., 2015)*. На слици се пореде два различита објекта O_1 и O_2 према истом критеријуму a. На вертикалној оси мери се степен поседовања неке особине a на интервалу $[0,1]$ и степен непоседовања те исте особине *a*. Укупна мера сличности је унија ове две компоненте. Другим речима, ИБА мера сличности посматра и имање и немање особине као део једног истог простора особина у коме је представљен неки објекат. Овај приступ сличности представља унију пресека поседовања особине и пресека непоседовања особине. Тиме се мери колико су објекти исти по некој особини али и колико су исти по немању те особине, односно сматра се да су они слични у погледу оба ова дела.



Слика 9. Графичка интерпретација мере сличности засноване на ИБА еквиваленцији

У реално-вредносној реализацији оба претходна израза за логичку еквиваленцију своде се на исти генерализован Булов полином односно ИБА еквиваленцију:

$$\begin{aligned}(a_1 \Leftrightarrow a_2)^{\otimes} &= \left((a_1 \Rightarrow a_2) \wedge (a_2 \Rightarrow a_1)\right)^{\otimes} = \\ &= \left((a_1 \wedge a_2) \vee (\neg a_1 \wedge \neg a_2)\right)^{\otimes} = \\ &= 1 - a_1 - a_2 + 2 \cdot \min(a_1, a_2)\end{aligned} \tag{4.5}$$

Занимљиво је приметити да је ИБА еквиваленција математички еквивалентна Лукашијевич би-импликацији иако ове две релације проистичу из различитих оквира. Међутим, ИБА еквиваленција је значајна са аспекта интерепретације.

Потреба за интерпретацијом је кључна при закључивању, које је у основи доношења одлука и давања савета *(Rissland, 2006).* Поред тога, сличности засновани на ИБА еквиваленцији се могу даље агрегисати са логичком агрегацијом која је такође заснована на ИБА *(Radojević, 2008a).*

4.3. Примена ИБА еквиваленције код закључивања на основу случајева

У овом делу се на илустративним примерима показује како се ИБА мера сличности може применити за одређивање сличности код проблема закључивању на основу случајева (*енг.* case-based reasoning). За истраживање проблема сличности код закључивања на основу случајева коришћени су подаци са тржишта некретнина у Србији, а полазни резултати су објављени у раду *(Poledica et al., 2010).*

Основни проблем је како наћи стан који највише задовољава потребе купца имајући у виду све станове који су расположиви на тржишту. За дату цену и карактеристике станова потребно је наћи један или више најсличнијих станова.

Закључивање на основу случајева је метод који се заснива на аналогији односно решава проблем тако што повезује претходно решен проблем или искуство са текућим нерешеним проблемом. При решавању новог проблема прво је потребно да се пронађу најсличнији случајеви из базе случаја. Овај корак представља један од најважнијих проблема код закључивања на основу случајева, а зависи од избора атрибута за репрезентацију случајева и од избора функције сличности *(Li & Ho, 2009).* Добра функција за проналажење случајева (*енг.* retrieval function) би требало да узме у обзир карактеристике станова који су важнији и да већу тежину додели случајевима који су најкориснији за посматрани домен. Ради успешног моделовања и изградње система закључивања на основу случајева, препоручује се интеграција доменског знања и субјективних оцена при конструисању функција за проналажења случајева у бази *(Richter & Aamodt, 2006*). Претходно наведени су основни мотиви за увођење логичке агрегације и мере сличности засноване на ИБА.

У овом истраживању, следеће карактеристике станова су узете у обзир: број соба, величина, локација, старост, спратовност, лифт, паркинг, грејање, балкон и цена. У Табели 2 наведен је извод из базе станова са нормализованим вредностима за свих десет атрибута.

Табела 2. Карактеристике станова - нормализоване вредности

A/a	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
A_1	0,50	0,40	1	0,20	0,60	1	0,5	0,2	0	0,32
A_7	0,65	0,60	0,50	1	0,60	1	1	1	0,25	0,28
...	...	...	...	...	...	...	...	...	...	...
A_{125}	0,75	0,71	0,50	0,20	0,35	1	0,50	0,40	1	0,40

У новом логичком приступу постоје два нивоа. Први се односи на мерење сличности одговорајућих особина станова коришћењем ИБА еквиваленције као мере сличности. Други ниво представља израчунавање глобалне сличности која се добија одговарајућом агрегацијом мера сличности станова, а која је у наставку илустрована на примеру.

Пример 1. У првом кораку се рачунају сличности *s* између одговарајућих атрибута за стан егземплар *A* и за сваки стан у бази појединачно. Претпостављамо да за наведени проблем постоји потреба да глобална сличност узима у обзир два парцијална захтева: 1) сличност цене и локације су важне ($S_1^{\otimes}$) и 2) сличност за величину стана је важна, али ако квадратура није велика онда се узима у обзир и број соба ($S_2^{\otimes}$). Логичке функције за ове захтеве се могу написати на следећи начин:

$$s_{10} \wedge s_3 \tag{4.6}$$

$$s_2 \vee \left(\neg s_2 \wedge s_3\right) \tag{4.7}$$

док су одговарајући ГБП и њихове вредносне реализације за производ као генерализовани производ следеће:

$$S_1^{\otimes} = s_{10} \otimes s_3 = 0,98 \quad (4.8)$$

$$S_2^{\otimes} = s_2 + (1 - s_2) \otimes s_1 - s_2 \otimes (1 - s_2) \otimes s_1 = 0,9523 \quad (4.9)$$

Глобална сличност се може изразити следећом псеудо-логичком функцијом где оба парцијална захтева имају једнаку тежину:

$$Agg^{\otimes}(s_1, s_2, s_3, s_{10}) = 0.5 \cdot S_1^{\otimes} + 0.5 \cdot S_2^{\otimes} = 0,9611 \quad (4.10)$$

У Табели 3 приказано је да је сличност пронађеног стана A_7 и егземплара A једнака 0,9661 и да је у складу са претходним захтевима у погледу критеријума за станове.

Табела 3. Задати (A) и пронађени стан (A_7) - нормализоване вредности

A/a	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
A	**0,50**	**0,45**	**0,50**	1	0,76	1	0,5	1	0,25	**0,18**
A_7	**0,40**	**0,40**	**0,50**	1	0,23	1	0	1	0,25	**0,16**

У раду (*Poledica et al., 2010*) наведено је неколико примера за различите агрегационе функције при одређивању глобалне мере сличности.

4.4. Поређење логичких мера сличности код проблема консензуса

У раду *(Poledica et al., 2015)* приказана је могућност примене мере сличности засноване на ИБА еквиваленцији за моделовање консензуса у групном одлучивању (*енг.* group decision making).

У основи модела налазе се две мере. Прва се односи на одређивање степена консензуса на нивоу групе, а друга на степен близине (*енг.* proximity) појединачног експертског мишљења мишљењу групе. Прво је потребно одредити матрице сличности експертских мишљења за сваку од алтернатива. Затим, на првом нивоу се израчунавају степен консензуса (CD, *енг.* consensus degree) и близине за сваку алтернативу појединачно. На основу тога се, на другом нивоу, израчунају укупни конзенсус и близина на нивоу групе. Детаљи о самом моделу и математичкој оправданости могу се пронаћи у *(Poledica et al., 2015)*, док је у овој докторској дисертацији циљ да се укратко прикажу могућности примене логичке мере сличности засноване на ИБА.

Примена консензус модела је илустрована на проблему избора најбољег пројекта у контексту одрживог развоја. У илустративном примеру четири пројекта су оцењивана од стране различитих експерата. Да бисмо узели у обзир све компоненте одрживог развоја, потребно је да експерти буду из области економије, друштва и екологије. Претпоставка је да су оцене експерата нормализоване на јединичном интервалу и приказане у Табели 4.

Табела 4. Пројекти оцењени од стране експерата - нормализоване вредности

Пројекат/експерт	E_1	E_2	E_3	E_4	Просек
$П_1$	0,94	0,87	0,73	0,69	0,808
$П_2$	0,37	0,30	0,25	0,19	0,278
$П_3$	0,89	0,79	0,20	0,30	0,545
$П_4$	0,70	0,56	0,20	0,30	0,440

Из Табеле 4 може се видети да је пројекат *П*1 високо оцењен од стране свих експерата, док је пројекат *П*2 оцењен као лош од стране свих експерата. Оно што је очекивано да ће ниво консензуса бити највећи за ове пројекте. Што се тиче осталих пројеката експертска мишљења се највише разликују за пројекат *П*3, те је интуитивно јасно да ће и ниво консензуса бити најнижи за овај пројекат.

У општем случају можемо рећи да се експерти слажу по питању неке алтернативе ако оба експерта оцењују алтернативу као добру или оба експерта оцењују алтернативу као лошу. Због тога, еквиваленција представља природан и интуитиван начин да се опише слагање односно консензус између експертских мишљења. Овде је приказано моделовање консензуса помоћу ИБА еквиваленције као једне од реализација реално-вредносне еквиваленције на јединичном интервалу. Тиме се добија „мека” мера која показује степен слагања између експертских мишљења. У Табели 5 су збирно приказане матрице сличности експертских мишљења за сваки од оцењених пројеката.

Табела 5. Степен консензуса за пројекте - ИБА мера сличности

	E_1, E_2	E_1, E_3	E_2, E_3	E_1, E_4	E_2, E_4	E_3, E_4
$П_1$	0,930	0,790	0,860	0,750	0,820	0,960
$П_2$	0,930	0,880	0,950	0,820	0,890	0,940
$П_3$	0,900	0,310	0,410	0,410	0,510	0,900
$П_4$	0,860	0,500	0,640	0,600	0,740	0,900

За сваки пројекат збирно је израчунат степен консензуса (Табела 6). Такође уведен је ранг ради лакшег поређења са осталим логичким мерама заснованим на минимум, производ и Лукашијевич фази би-импликацији, које су детаљно описане у раду (*Poledica et al., 2015*).

Табела 6. Степен консензуса и ранг за сваку од алтернатива

	CD_M	Ранг$_M$	CD_P	Ранг$_P$	CD_{IBA}	Ранг$_{IBA}$
$П_1$	0,733	**1**	0,836	**1**	0,852	2
$П_2$	0,228	4	0,704	2	0,902	**1**
$П_3$	0,332	2	0,458	4	0,573	4
$П_4$	0,293	3	0,512	3	0,707	3

Степен слагања између свих експерата на свим пројектима представља укупни групни консензус и добија се агрегацијом консензуса за сваки од пројеката на претходном нивоу. У овом случају износи $CD_{IBA} = 0,758$.

На основу резултата (Табела 6) закључујемо да ИБА еквиваленција може успешно да моделује консензус и то у складу са опште прихваћеном перцепцијом консензуса која је горе наведена. Другим речима, ИБА еквиваленција препознаје као сличне и случајеве када оба експерта дају добру оцену алтернативи, али и случај када се оба експерта слажу да је алтернатива лоша (пројекат *П*1 и *П*2). То није случај код већине логичких мера сличности нпр. CD_M не узима у обзир степен сличности према непоседовању неке особине, док се код CD_P обухвата само делимично *(Poledica et al., 2015).* Такође, CD_{IBA} једнако третира и сличности малих и сличности великих вредности, док CD_P фаворизује сличности великих вредности, што није адекватно код моделовања консензуса *(Poledica et al., 2015).* Такође је показано да Лукашијевич фази би-импликација даје исту вредност сличности као ИБА еквиваленција. Међутим, предност ИБА еквиваленције се огледа у њеној јасној мотивацији и могућношћу интерпретације што има велики значај у домену примене *(Poledica et al., 2015).*

Остали резултати добијени за мере близине појединачног експерта групном мишљењу налазе се у раду *(Poledica et al., 2015).*

5. ПАРАМЕТРИЗАЦИЈА ИБА МЕРЕ СЛИЧНОСТИ

Према претходно наведеној мери сличности заснованој на ИБА еквиваленцији (једначина 4.3), сличност између два објекта обухвата степен поседовања и степен непоседовања неке особине. Претпоставка је да је у претходном изразу функција минимума одговарајућа да одреди колико су два објекта иста према поседовању неке особине. У ИБА минимум се користи за особине које имају исту/сличну природу. Према *Radojević (2008a)* минимум је једина могућа/адекватна, јер је конзистентно поређење објеката могуће само према истој особини. У овом поглављу најпре се претпоставља да се део израза који се односи на непоседовање особине, $1-x-y+\min(x,y)$ може подешавати. Увођењем параметра, степен непоседовања особине може се прилагодити одређеном проблему односно скупу података. Параметризацијом се ИБА мера сличности уопштава, јер се у зависности од вредности параметра може свести на постојеће мере сличности, што ће се показати у одељку 5.1. У наставку (одељак 5.2) је приказана двопараметарска ИБА мера сличности код које се, осим израза који се односи на непоседовање особине, параметром подешава и део који се односи на поседовање особине.

5.1. Параметарска ИБА мере сличности - са једним параметром

Када се у израз за меру сличности засновану на ИБА еквиваленцији уведе параметар уз део који се односи на непоседовање особине добијамо следећи израз:

$$s_{IBA}^{\alpha}(x,y) = \alpha \cdot \left[1-x-y+\min(x,y)\right] + \min(x,y) \tag{5.1}$$

где параметар α узима вредности са јединичног интервала $[0,1]$. На тај начин вредност параметра α одређује значај дела мере који се односи на непоседовање особине. У случају да параметар α има вредности са крајева јединичног интервала, ИБА еквиваленција се своди на претходно приказане непараметарске мере сличности.

Стога се параметарска ИБА еквиваленција може приказати и на следећи начин::

$$s_{IBA}^{\alpha}(x,y)=\begin{cases}\min(x,y), & \alpha=0\\ \alpha\cdot\left[1-x-y+\min(x,y)\right]+\min(x,y), & \alpha=(0,1)\\ 1-x-y+2\cdot\min(x,y), & \alpha=1\end{cases} \tag{5.2}$$

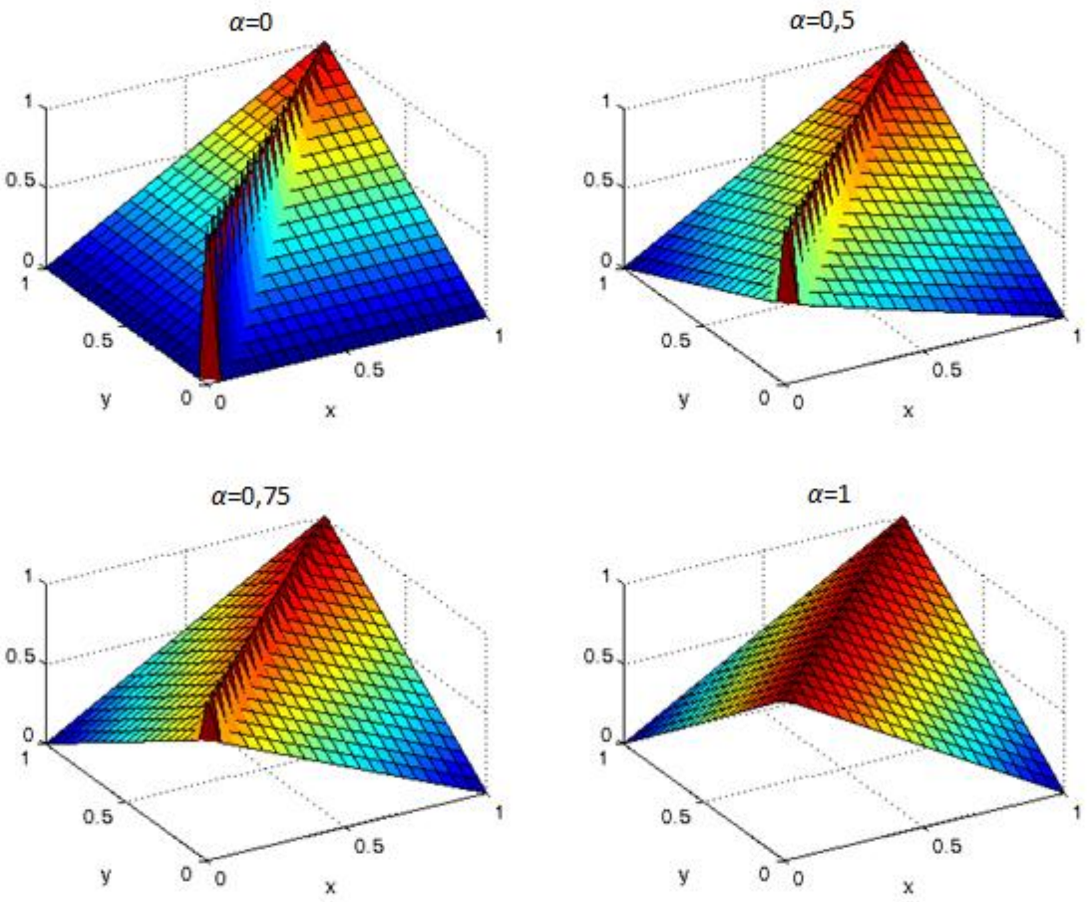
Међутим, битно је да се посебно узме у обзир случај када је $x=y$ јер тада сличност мора бити 1 будући да је објекат најсличнији са самим собом. Погоднији облик параметарске функције се може записати преко фази би-импликација заснованих на познатим *t*-нормама. Другим речима за $\alpha=0$, $s_{IBA}^{\alpha}(x,y)$ се своди на сличност засновану на минимум би-импликацији, за $\alpha=1$, $s_{IBA}^{\alpha}(x,y)$ постаје сличност заснована на ИБА еквиваленцији односно Лукашијевич би-импликацији, а за $\alpha=(0,1)$, $s_{IBA}^{\alpha}(x,y)$ је израз за параметарску ИБА еквиваленцију која обухвата и меру сличности засновану на производ би-импликацији:

$$S_{IBA}^{\alpha}(x,y)=\begin{cases}1, & x=y,\\ S_M(x,y), & \alpha=0,\ x\neq y,\\ \alpha\cdot\left[1-x-y+\min(x,y)\right]+\min(x,y)\supseteq S_P(x,y), & \alpha\in(0,1),\ x\neq y,\\ S_{IBA}(x,y)=S_L(x,y), & \alpha=1,\ x\neq y.\end{cases} \tag{5.3}$$

Тиме се показује да се параметарска ИБА мера сличности креће у следећем интервалу:

$$S_M(x,y)\leq S_{IBA}^{\alpha}(x,y)\leq S_L(x,y) \tag{5.4}$$

Параметарска ИБА еквиваленција за карактеристичне вредности параметра приказана је на Слици 10. Овде се јасно види да се неједнако третирају сличности малих и сличности великих вредности на јединичном интервалу, осим када је вредност параметра 1 односно код стандардне, непараметарске ИБА еквиваленције. Основне разлике између логичких мера познатих би-импликација већ су детаљно објашњене у претходном поглављу 4.4 на проблему консензуса.



Слика 10. Параметарска ИБА еквиваленција за различите вредности параметра

На примеру ће се показати како се параметарска ИБА еквиваленција $S_{IBA}^{\alpha}(x,y)$ може свести на производ би-импликацију $S_P(x,y)$ за одређену вредност параметра α на интервалу $(0,1)$.

Пример 2. Ако је $x=0,5,\ y=0,7$ онда је:

$$\begin{aligned} &S_M(x,y)=0,5,\\ &S_P(x,y)=0,714,\\ &S_{IBA}(x,y)=S_L(x,y)=0,8 \end{aligned} \tag{5.5}$$

Израз за параметарску ИБА еквиваленцију је:

$$S_{IBA}^{\alpha}(x,y)=(1-0,5-0,7+0,5)\cdot\alpha+0,5=0,3\cdot\alpha+0,5 \tag{5.6}$$

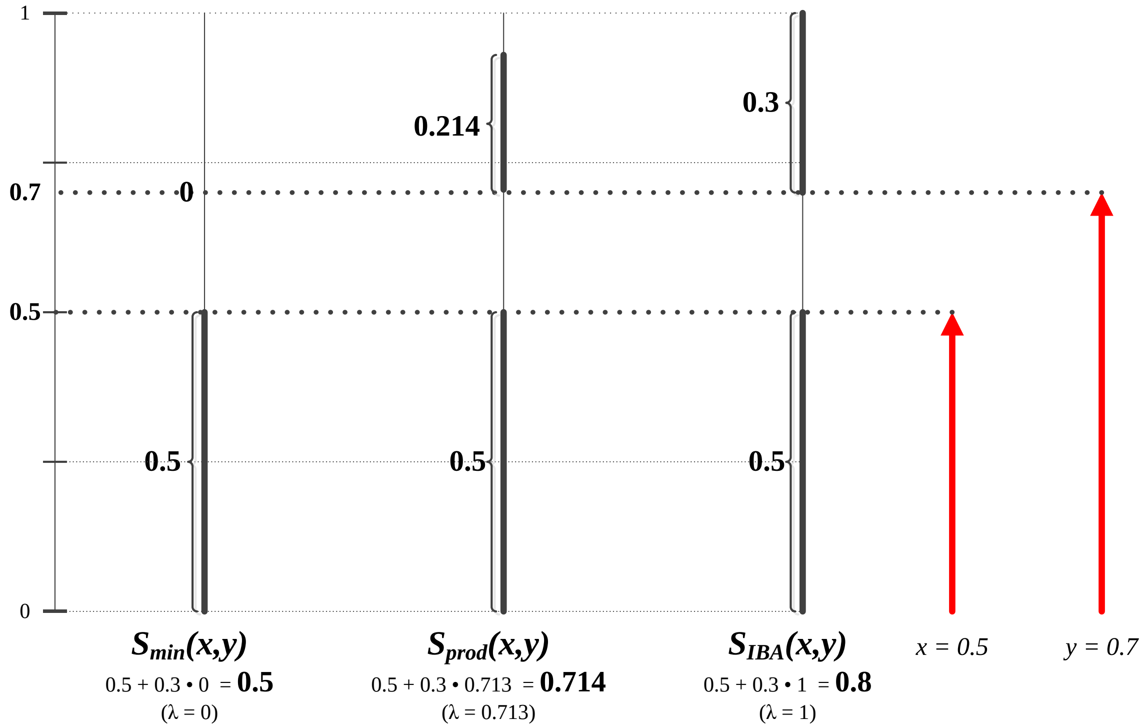
За вредност параметра $\alpha = 0,713$ добиће се да је $S_{IBA}^{\alpha}(x,y) = S_P(x,y)$.

$$\alpha = 0, \quad S_{IBA}^{\alpha}(x,y) = S_M(x,y) = 0,3 \cdot \alpha + 0,5 = 0,5 \tag{5.7}$$

$$\alpha = 1, \quad S_{IBA}^{\alpha}(x,y) = S_{IBA}(x,y) = 0,3 \cdot \alpha + 0,5 = 0,8 \tag{5.8}$$

$$\alpha = 0,713, \quad S_{IBA}^{\alpha}(x,y) = S_{prod}(x,y) = 0,3 \cdot \alpha + 0,5 = 0,214 + 0,5 = 0,714 \tag{5.9}$$

Претходни пример се може илустровати и графички (Слика 11).



Слика 11. Пример свођења параметарске ИБА евиваленције на друге мере сличности

Такође, могуће је и одредити вредност параметра α за који важи да је $S_{IBA}^{\alpha}(x,y) = S_P(x,y)$:

$$\begin{aligned}
S_{IBA}^{\alpha}(x,y) &= \left[1 - x - y + \min(x,y)\right] \cdot \alpha + \min(x,y) = S_P(x,y) \\
S_{IBA}^{\alpha}(x,y) &= \left[S_{IBA}(x,y) - S_M(x,y)\right] \cdot \alpha + S_M(x,y) = S_P(x,y) \\
\alpha &= \frac{S_P(x,y) - S_M(x,y)}{S_{IBA}(x,y) - S_M(x,y)}
\end{aligned} \tag{5.10}$$

Према једначини 5.10, параметар се добија као нормализована вредност на јединичном интервалу где је минимум S_M, а максимум S_{IBA} односно за крајње вредности параметарске ИБА еквиваленције $S_{IBA}^{\alpha}(x,y)$ за $\alpha=0$ и $\alpha=1$.

Код сложенијих параметарских модела често се дешава да модели нису довољно транспарентни и интерпретабилни. Значај овог модела је у томе што постоји јасна мотивација и графичко објашњење за увођење параметра. Математичка оправданост предложеног модела показује се провером својстава мере сличности.

Раније је показано да минимум, производ, Лукашијевич би-импликација и ИБА еквиваленција задовољавају услове за меру сличности *(Теорема 1, 2, 3 и 4)*. Може се претпоставити да ће и параметарска ИБА еквиваленција, која се за различите вредности параметара своди на претходне мере, такође задовољити услове за меру сличности. У наставку ће се то и експлицитно показати провером својстава за меру сличности.

Теорема 5. Нека је $S_{IBA}^{\alpha}:[0,1]^2 \to [0,1]$ релација еквиваленције у ИБА са параметром $\alpha \in [0,1]$. За било које $x,y \in [0,1]$, $S_{IBA}^{\alpha}(x,y)$ је мера сличности.

Доказ: Потребно је показати да S_{IBA}^{α} задовољава својства дата у *Дефиницији 8*:

Посматраћемо три случаја:

1. За $\alpha=0$, $S_{IBA}^{\alpha}=S_M$, S_M испуњава услове за меру сличности што је показано у *Теореми 1*.
2. За $\alpha=1$, $S_{IBA}^{\alpha}=S_L=S_{IBA}$, S_L и S_{IBA} испуњавају услове за меру сличности што је показано у *Теореми 3* и *4*.
3. За $\alpha \in (0,1)$ својства за меру сличности се проверавају:

Ненегативност: $S_{IBA}^{\alpha}(x,y) \geq 0$

$$S_{IBA}^{\alpha}(x,y)=\left[1-x-y+\min(x,y)\right]\cdot\alpha+\min(x,y)=$$
$$=\left[1-x-y+2\cdot\min(x,y)\right]\cdot\alpha+\min(x,y)-\alpha\cdot\min(x,y)=$$
$$=S_{IBA}(x,y)\cdot\alpha+\min(x,y)-\alpha\cdot\min(x,y)=S_{IBA}(x,y)\cdot\alpha+(1-\alpha)\cdot\min(x,y)\geq 0$$

Већ је показано у *Теореми 4* да $S_{IBA}^{\alpha}(x,y)\in[0,1]$ и како $(1-\alpha)\in(0,1)$, следи да израз $S_{IBA}(x,y)\cdot\alpha+(1-\alpha)\cdot\min(x,y)\geq 0$ односно $S_{IBA}^{\alpha}(x,y)$ испуњава услов ненегативности.

Симетричност: $S_{IBA}^{\alpha}(x,y)=S_{IBA}^{\alpha}(y,x)$

$$S_{IBA}^{\alpha}(x,y)=\left[1-x-y+\min(x,y)\right]\cdot\alpha+\min(x,y)=$$
$$=\left[1-y-x+\min(y,x)\right]\cdot\alpha+\min(y,x)=S_{IBA}^{\alpha}(y,x)$$

услед комутативности функције минимума $\min(x,y)=\min(y,x)$.

Ограниченост са горње стране: $S_{IBA}^{\alpha}(x,y)\leq S_{IBA}^{\alpha}(x,x)$ са једнакошћу ако и само ако је $x=y$

$$\left[1-x-y+\min(x,y)\right]\cdot\alpha+\min(x,y)<\left[1-x-y+\min(x,y)\right]+\min(x,y)=S_{IBA}(x,y)$$

пошто је $\alpha<1$. Како је за $S_{IBA}(x,y)$ показана ограниченост са горње стране и $S_{IBA}^{\alpha}(x,x)=1$, ограниченост са горње стране важи и за $S_{IBA}^{\alpha}(x,y),\alpha\in(0,1)$.

5.2. Параметарска ИБА мере сличности - са два параметра

Полазећи од претпоставке да део израза који се односи на имање особине можемо додатно да прилагодимо проблему/подацима уводимо још један параметар. У том случају мера сличности заснована на ИБА еквиваленцији садржи два независна параметра и добијамо следећи облик:

$$s_{IBA}^{\alpha,\beta}(x,y)=\alpha\cdot\left[1-x-y+\min(x,y)\right]+\beta\cdot\min(x,y) \tag{5.11}$$

при чему је објекат са самим собом најсличнији и у том случају сличност $s_{IBA}^{\alpha,\beta}(x,y)$ има вредност 1:

$$s_{IBA}^{\alpha,\beta}(x,y)=\begin{cases}1, & x=y,\\ \alpha\cdot\left[1-x-y+\min(x,y)\right]+\beta\cdot\min(x,y), \quad \alpha,\beta\in[0,1], & x\neq y.\end{cases} \tag{5.12}$$

Теорема 6. Нека је $S_{IBA}^{\alpha,\beta}:[0,1]^2 \to [0,1]$ релација еквиваленције у ИБА са параметрима $\alpha \in [0,1], \beta \in [0,1]$. За било које $x, y \in [0,1]$, $S_{IBA}^{\alpha,\beta}(x,y)$ је мера сличности.

Доказ: Потребно је показати да $S_{IBA}^{\alpha,\beta}$ задовољава својства дата у *Дефиницији 8*:

Ненегативност: $S_{IBA}^{\alpha,\beta}(x,y) \geq 0$

$$S_{IBA}^{\alpha,\beta}(x,y) = \alpha \cdot [1-x-y+\min(x,y)] + \beta \cdot \min(x,y) \geq$$
$$\geq \min(\alpha,\beta) \cdot [1-x-y+\min(x,y)+\min(x,y)] =$$
$$= \min(\alpha,\beta) \cdot S_{IBA}(x,y) \geq 0$$

пошто је $S_{IBA} \in [0,1]$ (показано у *Теореми 4*) и $\alpha \in [0,1], \beta \in [0,1]$. Тиме је показано да важи ненегативности.

Симетричност: $S_{IBA}^{\alpha,\beta}(x,y) = S_{IBA}^{\alpha,\beta}(y,x)$

$$S_{IBA}^{\alpha,\beta}(x,y) = \alpha \cdot [1-x-y+\min(x,y)] + \beta \cdot \min(x,y) =$$
$$= \alpha \cdot [1-y-x+\min(y,x)] + \beta \cdot \min(y,x) = S_{IBA}^{\alpha,\beta}(y,x)$$

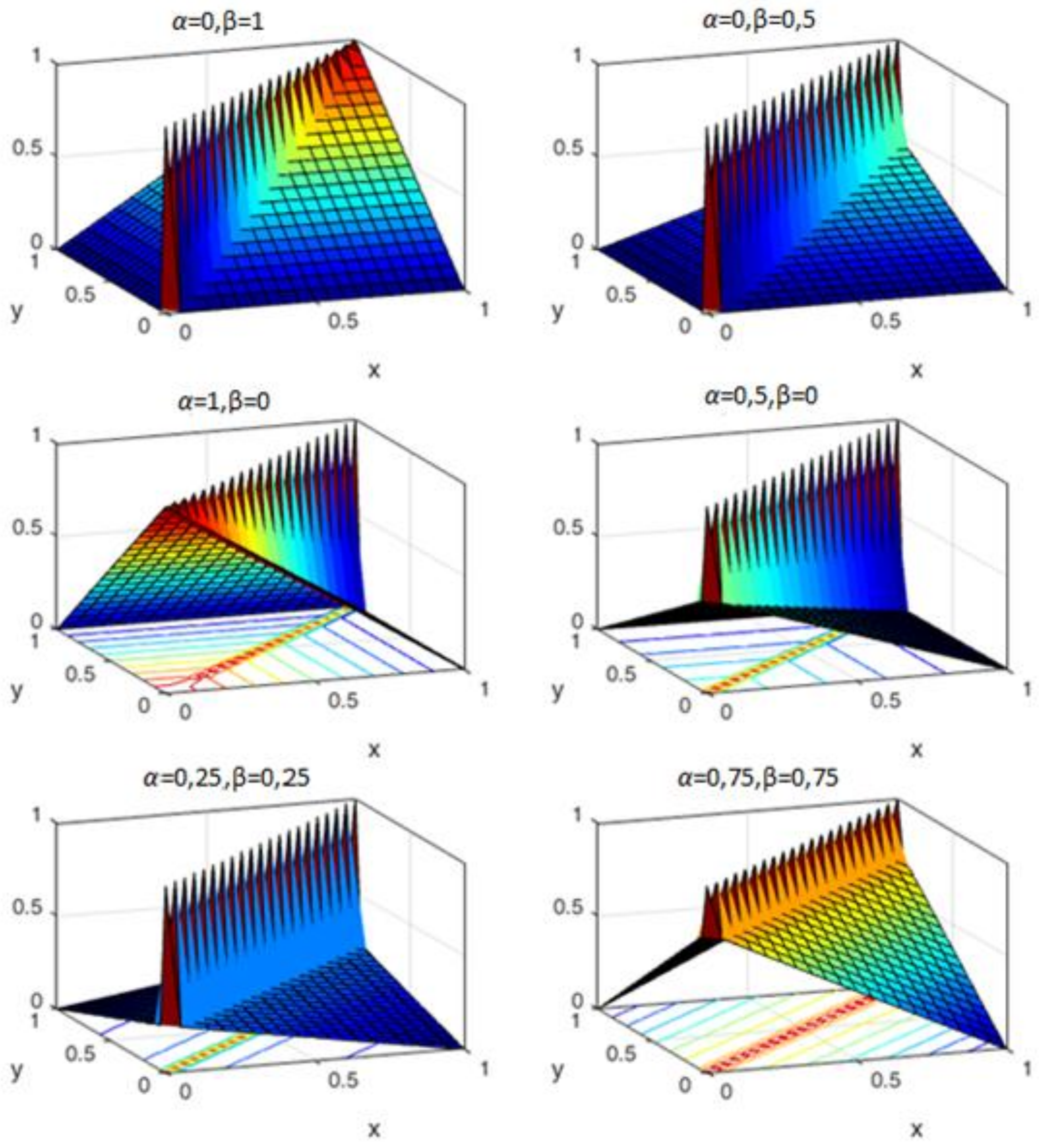
услед комутативности функције минимума $\min(x,y) = \min(y,x)$.

Ограниченост са горње стране: $S_{IBA}^{\alpha,\beta}(x,y) \leq S_{IBA}^{\alpha,\beta}(x,x)$ са једнакошћу ако и само ако је $x = y$.

Проверава се да ли је $S_{IBA}^{\alpha,\beta}(y,x) < S_{IBA}^{\alpha,\beta}(x,x)$ за $x \neq y$. Важи да је $S_{IBA}^{\alpha,\beta}(x,x) = 1$ без обзира на вредности параметара. Односно треба да се провери да ли је $S_{IBA}^{\alpha,\beta}(x,y) < 1$ за $x \neq y$.

$S_{IBA}^{\alpha,\beta}(x,y) \leq S_{IBA}(x,y) < S_{IBA}(x,x) = 1$ следи да је $S_{IBA}^{\alpha,\beta}(x,y) < 1$.

Двопараметарска ИБА еквиваленција за карактеристичне вредности параметра приказана је на Слици 12. Као и у случају једнопараметарске мере јасно се види да се неједнако третирају сличности малих и сличности великих вредности на јединичном интервалу у случају да су вредности параметара α и β различите. У случају $\alpha > \beta$, сличност малих вредности има значајно већу важност од сличности великих вредности. Другим речима, део израза који се односи на сличност непоседовања особине има већи значај. Обратно, ако је $\alpha < \beta$ добијене површи су сличне онима добијеним за непараметарске логичке мере сличности и једнопараметарску ИБА меру - сличности малих блиских вредности су мање него сличности великих блиских вредности. Када су вредности параметара α и β међусобно једнаке, површи су симетричне у односу на дијагоналу (као у случају ИБА мере сличности). Тада се сличности малих вредности и сличности великих вредности једнако третирају. У складу са једначином 5.12, сличност између истих вредности $x = y$ је увек једнака 1. Уколико се вредности x и y макар мало разликују, њихова сличност се смањује зависно од вредности параметара.



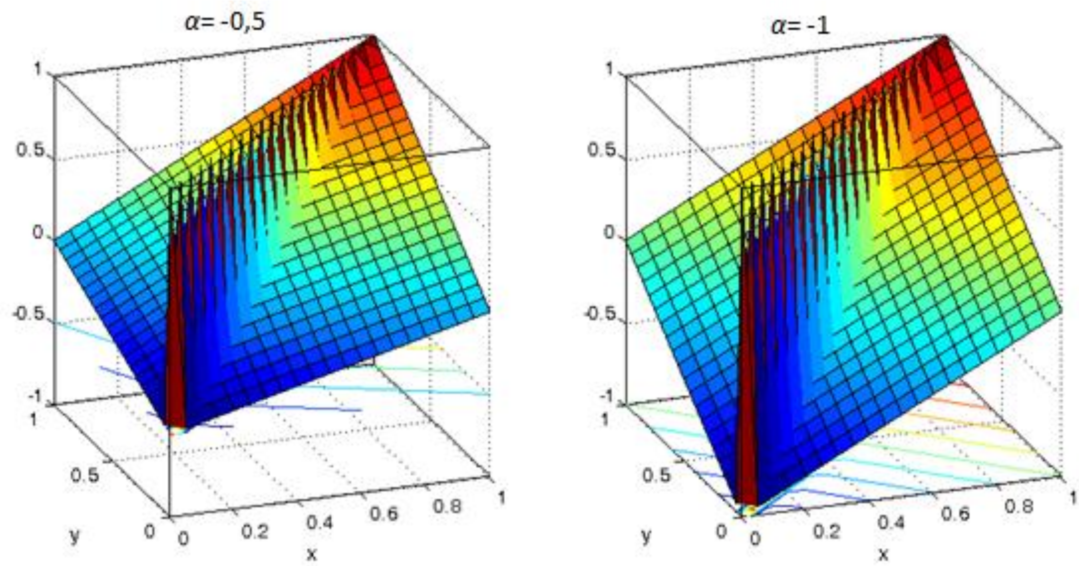
Слика 12. Параметарска ИБА мера сличности са два параметра

5.3. Могућности проширења приступа

До сада се при параметризацији ИБА еквиваленције претпостављало да параметри припадају јединичном интервалу $\left[0,1\right]$. Уколико се уведе да вредност параметра буде од -1 до 1, релација којом се мери сличност има следећи облик:

$$S_{IBA}^{\alpha[-1,1]}\left(x,y\right)=\begin{cases}1, & x=y,\\ S_M\left(x,y\right),\alpha=0, & x\neq y,\\ \alpha\cdot\left[1-x-y+\min\left(x,y\right)\right]+\min\left(x,y\right)\supseteq S_P\left(x,y\right), & \alpha\in\left(0,1\right),\ x\neq y,\\ S_{IBA}\left(x,y\right)=S_L\left(x,y\right), & \alpha=1,\ x\neq y,\\ \alpha\cdot\left[1-x-y+\min\left(x,y\right)\right]+\min\left(x,y\right), & \alpha\in\left(-1,0\right),\ x\neq y,\\ x+y-1, & \alpha=-1,\ x\neq y.\end{cases}\qquad(5.13)$$

Међутим, тиме се губе нека својства мера сличности (нпр. својство ненегативности) и излази се из теоријског оквира интерполативне Булове алгебре. Са друге стране, отвара се могућност за потенцијално другачије тумачење параметарске мере сличности. Идеја је да се овај приступ размотри у будућим истраживањима, односно да се обезбеди одговарајућа теоријска потпора. На Слици 13 приказани су примери параметарске мере сличности када су вредности параметра негативне.



Слика 13. Параметарска ИБА мера сличности

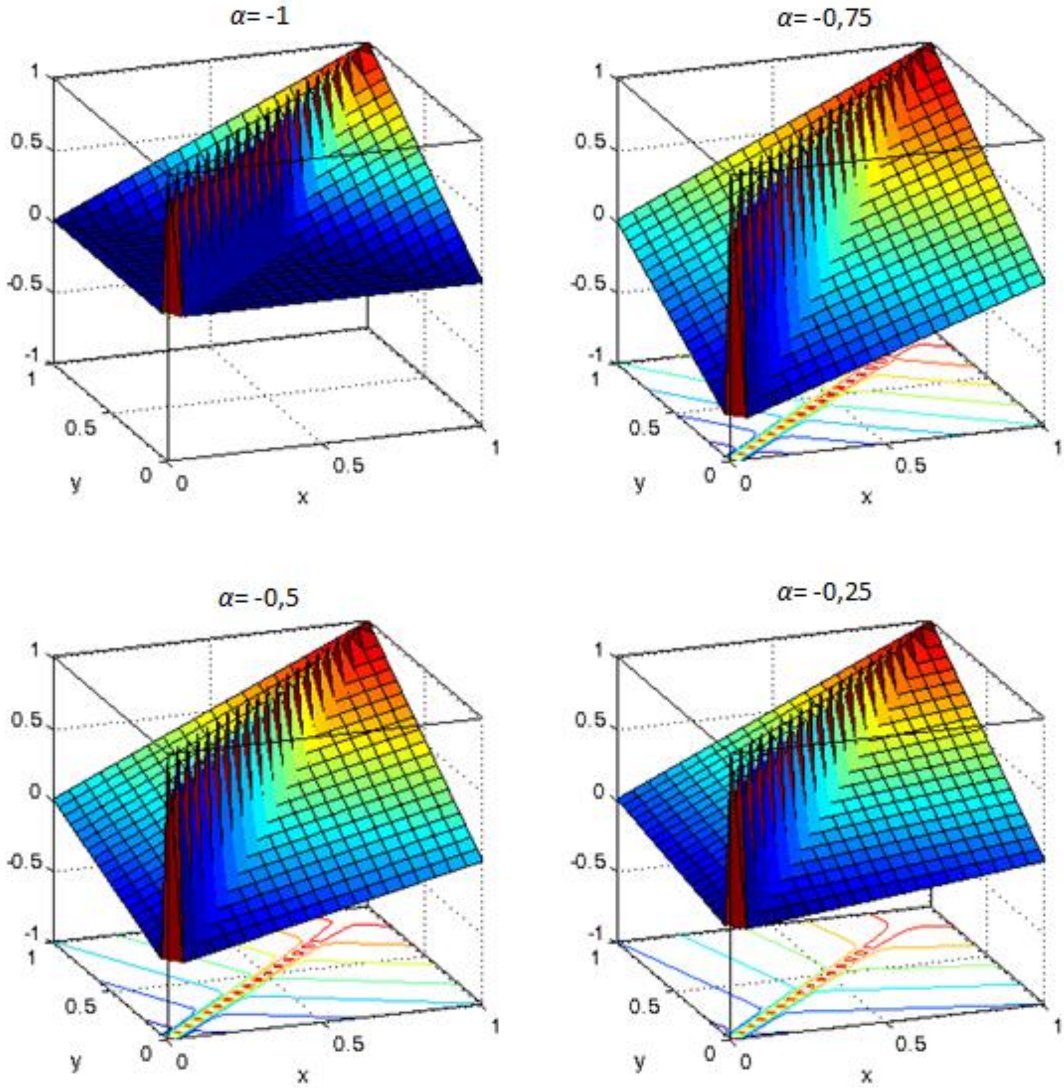
Једно од могућих решења је да се одговарајућом трансформацијом интервал [-1,1] сведа на јединични интервал $[0,1]$ *(Radojević, 2015)*.

Са друге стране, увођењем функције максимума негативне вредности за сличност се не узимају у обзир већ се своде на нулу уколико се појаве. Стога, могуће је сводити вредности на $[0,1]$ интервал на следећи начин:

$$S_{IBA}^{\alpha[-1,1]}(x,y)=\begin{cases}1, & x=y,\\ S_M(x,y), \alpha=0, & x\neq y,\\ \alpha\cdot\left[1-x-y+\min(x,y)\right]+\min(x,y)\supseteq S_P(x,y), & \alpha\in(0,1),\ x\neq y,\\ S_{IBA}(x,y)=S_L(x,y), & \alpha=1,\ x\neq y,\\ \max\left(\alpha\cdot\left[1-x-y+\min(x,y)\right]+\min(x,y),0\right), & \alpha\in(-1,0),\ x\neq y,\\ \max(x+y-1,0)=T_L(x,y), & \alpha=-1,\ x\neq y.\end{cases} \tag{5.14}$$

Занимљиво је приметити да се за вредност параметра -1 добија израз за Лукашијевич *t*-норму. Посматрајући израз може се уочити да су овде практично значајне само вредности када су $x+y>1$, јер су остале сличности нула. Тиме се искључују/занемарују случајеви када су обе вредности мале. Природно се намеће питање када се ова мера може показати као одговарајућа. Одговор на ово питање захтева додатна емпиријска и теоријска истраживања за модификовану меру поређења и биће предмет будућих истраживања.

На Слици 14. су приказани случајеви модификоване параметарске мере за различите негативне вредности параметра.



Слика 14. Модификована параметарска ИБА мера сличности

6. КЛАСИФИКАЦИЈА

6.1. Проблем класификације

Класификација је један од најчешћих задатака који се јавља при доношењу одлука људи и/или машина. При решавању проблема класификације истовремено се захтева и тачност и разумевање разлога за доношење неке одлуке *(Koloseni, 2015).* Неки примери класификације у реалном животу су (*Larose, 2014)*: одређивање да ли трансакција на кредитној картици спада у превару или не, оцена апликације за хипотекарни кредит као потенцијално добар или лош кредитни ризик, успостављање дијагнозе да ли је присутна нека болест код пацијента, одређивање да ли је тестамент написан од стране преминуле особе или је фалсификован од стране неке друге особе, детекција мрежне крађе идентитета (*енг*. phishing), итд.

Проблем класификације се јавља у различитим доменима примене. У финансијама класификационе технике се користе за предвиђање кретања тржишта *(e.g. Huang et al., 2005). Hand & Henley (1997)* су дали преглед статистичких метода које се користе за одобравања кредита, док су *Huang et al. (2004)* направили упоредно истраживање неуронских мрежа и машине потпорних вектора (*енг*. support vector machines) за анализу кредитног рејтинга. Преглед и поређење алгоритама за класификацију у домену медицине дао је *(Harper, 2005).* Алгоритми који се користе за класификацију текста могу се пронаћи у *(Aggarwal & Zhai 2012)*, а за проблем класификације слика у *(Lu & Weng, 2007).* Чест проблем који се јавља је и класификација интернет саобраћаја, а детаљан преглед је дат у *(Nguyen & Armitage, 2008).*

Основна идеја неке класификационе технике је да систематично изгради класификациони модел од улазног скупа података. Да би идентификовао модел који најбоље описује везу између променљивих у улазним подацима и класе као излазни податак, класификатор користи алгоритам за учење. Модел генерисан од стране алгоритма за учење би требало да добро описује улазне податке и да коректно предвиђа класу за нове инстанце *(Koloseni, 2015).* Другим речима, класификатор би требало да има добру способност генерализације. *Larose (2014)* је дао приказ за оптимални ниво комплексности модела који се налази у тачки минималне грешке на скупу података за тестирање. Како расте комплексност модела смањује се и стопа грешке на скупу за тренирања и на скупу за тестирање. Са растом сложености модела грешка тренирања наставља да опада монотоно, док грешка тестирања убрзо успорава пад и почиње да расте. То се дешава услед тога што је модел „преучио" односно изгубила се способност генерализације модела на новим подацима.

Класификација спада у проблеме надгледаног учења. Детаљан преглед класификационих техника са надгледаним учењем дат је код *(Kotsiantis, 2007)*. Код класификације и регресије излаз односно циљна променљива (лабела класе или реални број) је позната и користи се при обучавању модела. Код ненадгледаног учења циљна променљива није идентификована већ алгоритам покушава да препозна правилности и структуру у оквиру свих променљивих. Овај начин учења се може користити код компресије података, препроцесирања, кластеровања, редукције димензија или визуелизације података *(Han et al., 2011; Larose, 2014)*.

Насупрот методама учења којима се конструише општи опис циљне функције за обезбеђен тренинг скуп података, постоји и учење засновано на инстанцама (*енг.* instance-based learning). Код ових метода се чува тренинг скуп, а генерализација се одлаже док не постоји захтев да се класификује нова инстанца. Сваки пут се за нову инстанцу испитује однос са претходно сачуваним инстанцама да би се одредила вредност циљне функције за нову инстанцу. Учење засновано на инстанцама се често назива и лењо учење (*енг.* lazy learning) јер се процесирање одлаже до тренутка класификације нове инстанце. Основна предност овог начина учења је тај што уместо да само једном процењује циљну функцију за цео простор инстанци, ова метода врши оцену локално и различито за сваку инстанцу коју класификује (*Mitchell, 1997)*. Најпознатије методе учења заснованог на инстанцама су *k*-најближих суседа, закључивање на основу случајева, итд.

За ово истраживање посебно су значајни класификатори који се заснивају на сличности. *Chen et al. (2009)* дали су опсежан преглед алгоритама који користе сличност при класификацији. Група класификатора који се заснивају на генерализованој Лукашијевич структури уведени су код (*Luukka et al., 2001; Luukka 2005; Luukka & Leppälampi, 2006*). *Saastamoinen et al. (2002b)* истражују да ли се резултати могу унапредити ако се класификатори дефинисани у оквиру генерализоване Лукашијевич структуре прошире по угледу на општу и кумулативну Минковски меру сличности. Основна претпоставка је да ће кумулативна мера дати боље резултате када су подаци високо корелисани и уређени. Резултати класификације су били добри али не као што је очекивано. За класификацију код медицинских података *Luukka & Leppälampi (2006)* користе генерализовану Лукашијевич структуру са генерализованом средњом вредношћу и тежинама. Овако формиран класификатор заснован на фази сличности, показао се као добар за дијагнозу у случају посматраних медицинских података. Предложени модел класификације поредили су са другим класификаторима као што су линеарна дискриминантна анализа, различита стабла одлучивања, једнослојне и вишеслојне неуронске мреже, и за већину података дао је боље резултате. Даља истраживања *(e.g. Luukka 2009; Luukka 2011)* су представљала покушај да се унапреде резултати класификације засноване на сличности препроцесирањем података односно коришћењем анализе главних компоненти, фази анализе главних компоненти и увођењем фази мера ентропије као алата за редукцију података. Разматрана је и могућност за увођење различитих оператора агрегације *(e.g. Luukka & Kurama, 2013)*. Прва истраживања на тему формализације процеса људске категоризације помоћу интерполативне Булове алгебре дата су код *(Dobrić et al., 2015)*.

У наставку ће бити описан алгоритам најближег прототипа и *k*-најближих суседа јер они представљају основу за формирање нових логичких класификатора који су засновани на сличности.

6.2. Класификација заснована на прототипу

Одређивање прототипова и закључивање на основу сличности са прототиповима налази се у основи људског размишљања. У психологији и когнитивном моделовању постоји велики број радова који су се бавили прототипском класификацијом *(Reed, 1972; Donald et al. 1973; Medin & Schaffer, 1978; Rosch, 1983)*.

Класификација заснована на најближем прототипу је један од најједноставнијих алгоритама у области препознавања патерна. Прототип се може посматрати као репрезентативни представник сваке класе. Објекту се додељује лабела класе на основу близине односно сличности објекта једном или више прототипова. Основни проблем који се јавља је како пронаћи односно идентификовати одговарајуће прототипове на којима се заснива класификација. Циљ је да се одабере скуп прототипова тако да најбоље представљају посматрану класу. У идеалном случају, прототип би требало да има способност да обухвати све варијације у оквиру класе, при томе узимајући у обзир како се посматрана класа разликује од осталих класа *(Koloseni, 2015)*.

Најпознатији типови класификатора засновани на прототипу су *k*-најближих суседа *(Fix & Hodges, 1951; Cover & Hart, 1967)* и *k*-средњих вредности (*енг. k*-means) *(Steinhaus, 1956; MacQueen, 1967)*. Први не врши апстракцију података већ користи све податке за обучавање. То је типичан пример алгоритма са лењим учењем са онолико прототипова колико има и инстанци у подацима за тренинг. Са друге стране, класификатор који се заснива на удаљености од средње вредности чува само средњу вредност класе односно један прототип за сваку класу *(Veenman & Reinders, 2005)*.

У оквиру овог истраживања користи се метод најближег центроида и *k*-најближих суседа. Метода најближег центроида је брз и једноставан метод за класификацију *(Friedman et al., 2001; Levner, 2005)*. Овај алгоритам подразумева да свакој класи одговара један центроид који представља центар класе. Центроид се израчунава као аритметичка средина. Некласификованим инстанцама скупа за тестирање додељује се класа која одговара прототипу који је најближи тој инстанци. Обично се за одређивање близине инстанце од центроида користи Еуклидско одстојање, али се може изабрати и нека друга мера у зависности од проблема.

Поред тога, могуће је конструисати центроид класе и на други начин. На пример, метода *Nearest Shrunken Centroid* је посебан алгоритам за селекцију атрибута код класификатора најближег центроида *(Friedman et al., 2001)*. Овај алгоритам има за циљ да помери прототипове класа ка општој средњој вредности (*Levner, 2005)*. Алгоритам *k*-најближих суседа биће детаљније објашњен у следећем одељку.

6.3. Метода *k*-најближих суседа

Метода *k*-најближих суседа (*k*-NN, *енг.* *k*-Nearest Neighbors) је један од најчешће коришћених алгоритама за класификацију пре свега због своје једноставности, а делом што је дуго присутан у науци *(Mitchell, 1997)*. Овај класификатор увели су *Fix & Hodges (1951)*, а затим су *Cover & Hart (1967)* представили теоријске резултате. Свеобухватан преглед истраживања која се баве методом најближих суседа може се пронаћи у *(Dasarathy, 1991)*.

Постоји неколико отворених питања приликом изградње класификатора помоћу *k*-NN *(Larose, 2014)*:

- Колико суседа треба узети у обзир, односно колика је вредност *k*?
- Како меримо растојање, односно сличност?
- Како комбинујемо информације када посматрамо више од једног суседа?
- Да ли све инстанце имају исту тежину или је потребно да неке вредности имају већи утицај у односу на друге?

Код алгоритма *k*-најближих суседа за мерење сличности користи се нека метрика, међу којима најчешће Еуклидско одстојање *(Larose, 2014)*. Када је познат метод за одређивање инстанци које су најсличније новој некласификованој инстанци, потребно је дефинисати како се ове сличне инстанце комбинују да би се донела одлука којој класи припада нова инстанца. Најчешћа комбинациона функција која се користи за додељивање класе је једноставно гласање без тежина.

Код гласања без тежина вредност *k* одређује колико се најсличнијих инстанци узима у обзир при класификацији нове инстанце. Овде је правило једноставно - једна инстанца један глас. Код тежинског гласања идеја је да сличније инстанце добију већу тежину од суседа који су нешто даљи. Тиме се смањује вероватноћа да дође до нерешеног гласања. Овде је утицај неке инстанце инверзно пропорционалан удаљености инстанце новој некласификованој инстанци. Гласови за инстанце се отежавају у складу са инверзним квадратом њихових дистанци. У случају да је удаљеност нула, бира се класа којој припада већина инстанци код којих је удаљеност од нове инстанце нула *(Larose, 2014).*

Код метода учења заснованих на инстанцама (нпр. *k*-NN или закључивање на основу случајева) веома је важно да постоји приступ богатој и репрезентативној бази података са што више различитих комбинација вредности променљивих. Посебно је важно да ређи случајеви класификације буду присутни у довољној мери да алгоритам не би предвиђао само оне случајеве који преовлађују у бази. Због тога је потребно да скуп података за класификацију буде балансиран. Један од начина да се изврши балансирање је да се смањи проценат инстанци који се чешће јављају у бази. Одржавање овакве базе података за брзи приступ може бити проблематично ако постоје ограничења у погледу главне меморије. Главна меморија се може попунити, а приступ помоћном складишту је обично спор. Због тога решење може бити да се у бази задрже само оне инстанце које су ближе граници класификације *(Larose, 2014).*

Одабир вредности *k* представља још једно отворено питање код алгоритма најближих суседа. Уколико се за *k* узме мала вредност онда постоји могућност да постојање изузетака или шума драстично утиче на резултате класификације. Са друге стране, уколико је *k* превише велико могуће је превидети локално понашање које може, а не мора бити од интереса за класификацију. Једно од решења је да се *k* одреди уз помоћ података, односно да се одабере *k* такво да даје најмању грешку класификације добијене за различите случајно одабране тренинг скупове података помоћу унакрсне валидације *(Larose, 2014).*

Детаљнија анализа мана и предности методе најближих суседа дата је у *(Mitchell, 1997).* Алгоритам *k*-најближих суседа са тежинским гласањем је високо ефективан индуктиван метод за решавање многих практичних проблема. Робустан је у погледу тренинг података са шумом и прилично ефективан када постоји довољно велики тренинг скуп података. Тежинским просеком *k* најближих суседа ублажава се утицај изолованих изузетака у тренинг скупу. Код примене *k*-NN-а удаљеност између инстанци рачуна се на основу свих атрибута инстанце. Међутим то није случај код неких других метода као што су стабла одлучивања која бирију подскуп атрибута инстанце при формирању хипотеза. Код одређених проблема само неки од атрибута могу бити релевантни за класификацију, па би у том случају код *k*-NN-а велики број ирелевантних атрибута утицао на одређивање удаљености. *k*-NN је посебно осетљив на проблем великих димензија (*енг.* curse of dimensionality). Један од начина да се превазиђе овај проблем је додавање тежина сваком атрибуту приликом одређивања удаљености између инстанци. Овај приступ одговара растезању оса у Еуклидском простору, где скраћивање оса одговара мање релевантним атрибутима, а продужавање оса значајнијим атрибутима. Коришћењем унакрсне валидације могуће је оценити тежине атрибута односно оптимизовати перформансе алгоритма. Драстичнија алтернатива би била да се из простора инстанци комплетно елиминишу најмање релевантни атрибути, што је еквивалентно као да смо неке тежине поставили на нулу. *Moore & Lee (1994)* су анализирали методе унакрсне валидације које су ефикасне за одабир релевантног подскупа атрибута код *k*-најближих суседа. Посебно су истраживали унакрсну валидацију са једним издвојеним елементом (*енг.* leave-one-out cross-validation) која је једноставна за имплементацију у случају *k*-NN-а. Takoђе, треба имати у виду да се повећањем степена слободе за редефинисање функције удаљености повећава и ризик функције да 'преучи', односно да има смањену способност генерализације *(Mitchell, 1997).*

Још једна практична тема за разматрање код *k*-NN алгоритма је ефикасно индексирање меморије. Потенцијална мана методе најближих суседа је да она не изграђује модел, већ се ослања на чување свих података из тренинг скупа. Ако је тренинг база података велика, потребно је значајно рачунарско време за процесирање сваког упита. Конкретно може захтевати *O(np)* операција за сваку инстанцу упита, обилазећи сваку од *n* инстанци тренинг података и извршавајући *p* операција за израчунавање удаљености до сваке инстанце. Рачунска ефикасност алгоритма најближих суседа описана је у *(Preparata & Shamos, 1985)*. Са аспекта меморије, алгоритам захтева да се сачува цео тренинг скуп величине *np*. Директна примена овог алгоритма није практична код апликација које укључују велике вредности за *n* и/или класификацију у реалном времену. Бројне методе су развијене да се убрза претрага и захтеви за меморијом код основног приступа као на пример гранање и ограничавање (*енг.* branch and bound) *(Hand et al., 2001)*. Такође, *k*-димензионо стабло претраживања (*енг. k*-d tree) *(Bentley 1975; Friedman et al. 1977)* је једна од метода за индексирање сачуваних података која помаже да се најближи суседи идентификују ефикасније. Овде су инстанце сачуване као листови стабла где се ближе инстанце чувају на истим или оближњим чворовима. Уколико се чуване тест инстанце организују у стабла претраживања у случају 1-NN, број поређења се може смањити на *O(log(n))*, где *n* представља податке за тестирање *(Han et al., 2014)*.

6.4. Препроцесирање података

У овом поглављу се објашњава важност препроцесирања података. Препроцесирање података је значајан корак при моделовању великог броја проблема јер се добре одлуке могу донети само на основу квалитетних података. Сирови подаци који се налазе у базама и складиштима података су често необрађени, некомплетни и имају шум. На пример, база података може садржати поља која су застарела или редудантна, затим може имати недостајуће вредности, изузетке, податке који нису у одговарајућој форми за примену жељених модела, неконзистенте вредности *(Larose, 2014)*. Технике препроцесирања података могу значајно допринети укупном квалитету података и самим тим побољшају тачност и ефикасност алгоритама класификације *(Han et al., 2011).*

Припрема односно препроцесирање података се најчешће врши у облику пречишћавања (*енг.* data cleaning), интеграције (*енг.* data integration), трансформације (*енг.* data transformation) и редукције података (*енг.* data reduction *(Han et al., 2011).* Пречишћавање података се примењује са циљем да се уклони шум и неконзистентност у подацима. Интеграција података спаја податке из више различитих извора у кохерентно складиште података. Такође, могуће је применити трансформацију података попут нормализација података, чиме се може унапредити тачност и ефикасност алгоритама који у себи садрже мерење дистанце/сличности. Техникама редукције података попут агрегације, елиминацијом редудантних карактеристика, кластеровањем и сл. могу се смањити димензије података. Ове технике се не искључују међусобно већ се могу комбиновати заједно. На пример, чишћење података може садржати и трансформације које се односе на исправљање некоректних података. Основни циљ је да се обезбеде што квалитетнији подаци, јер лоши подаци на улазу у модел неминовно дају лош излаз из модела (*енг.* garbage in, garbage out). Процењено је да сама припрема података односи 60% времена и труда уложеног у целокупни процес моделовања неког проблема *(Pyle, 1999).*

Веома је важно анализирати податке и њихове карактеристике методама и техникама дескриптивне статистике. Ове технике нам помажу да детектујемо шум, изузетке и разумемо расподелу података. Поред тога, значајно је да се избегне коришћење корелисаних променљивих као улаз у неки модел. У најбољем случају корелисане променљиве ће појачати утицај једне компоненте у подацима. У најгорем случају, корелисане променљиве утичу да модел постане нестабилан и да производи непоуздане резултате. Једноставно решење би било да се елиминишу редудантне променљиве, чиме се редукује и димензионалност простора решења. Поред тога, потребно је испитати и постојање веза и односа између више променљивих. Графичке методе као што је дво- или тродимензионални дијаграм растурања могу се користити да се утврди да ли у нумеричким подацима постоји интеракција, патерн или тренд, и да ли се у подацима издвајају кластери или неки изузеци *(Larose, 2014)*.

Промењљиве често имају веома различите опсеге вредности. Поред опсега и начин на који су вредности промењљивих распоређене такође може бити значајан. Код неких алата за моделовање разлика у опсегу вредности доводи до тога да промењљиве са већим опсегом имају неодговарајући утицај на резултате. Због тога је потребно нормализовати податке како би се стандардизовала скала утицаја коју свака промењљива има на резултате *(Larose, 2014)*. Могуће је вршити нормализацију опсега промењљивих и нормализацију расподеле промењљивих *(Pyle, 1999)*.

6.4.1. Нормализација података

Код нормализације подаци се скалирају тако да њихове вредности буду у неком специфичном опсегу. Најчешће су то интервали од -1 до 1 и од 0 до 1. Нормализација података је посебно корисна за алгоритме класификације као што су неуронске мреже или алгоритми засновани на дистанци - најближи суседи и кластеровање. Код неуронских мрежа нормализација помаже да се убрза фаза учења, док код методе најближих суседа спречава превелики утицај промењљивих које су почетно припадале великим интервалима у односу на оне са почетно малим интервалима *(Han et al., 2011)*.

Постоје многобројне методе које се користе за нормализацију података. Основе најкоришћенијих метода су дате у наставку, при чему X означава оригиналну вредност, а X^* нормализовану вредност.

Мин-макс нормализација је линеарна траснформација оригиналних података. Она мапира вредност X у вредност X^* у опсегу $\left[novi\_\min(X), novi\_\max(X)\right]$ на следећи начин:

$$X^* = novi\_\min(X) + \frac{X - \min(X)}{\max(X) - \min(X)}\left(novi\_\max(X) - novi\_\min(X)\right) \quad (6.1)$$

где су $\min(X)$ и $\max(X)$ минимум и максимум вредности X. Мин-макс нормализација чува зависност између вредности у оригиналним подацима. Доћиће до грешке уколико нова улазна вредност буде ван опсега оригиналних података *(Han et al., 2011).*

Код линеарне нормализације, вредност X се може нормализовати и на следећи начин:

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (6.2)$$

Овде се вредност X скалира у оквиру јединичног интервала $[0,1]$. Мин-макс нормализација има за циљ да скалира променљиву у складу са разликом вредност те променљиве у односу на минималну вредност. Из формуле је јасно да ће минимум за неку променљиву имати нормализовану вредност 0, а максимум нормализовану вредност 1. Такође, можемо очекивати да вредности близу центра дистрибуције имају нормализовану вредност близу 0.5 *(Larose, 2014).*

Линеарно скалирање има смисла када су познате вредности за максимум и минимум. Међутим, приликом препроцесирања података максимум и минимум су познате за узорак али не и за популацију. Због тога може доћи до грешке уколико нова вредност излази ван опсега узорка. За реалну примену предлаже се модификација линеарне трансформације помоћу меког-макс скалирања (*енг.* softmax scaling) које има особину да "меко" тежи максималној вредности, али никада не достиже максималну вредност. Пример такве функције је логистичка функција *(Pyle, 1999).*

***z*-скор стандардизација (нула-средња вредност нормализација)** је широко распространјена у статистичкој анализи. Код *z*-скор стандардизације вредности X се нормализују одузимањем средње вредности и скалирањем разлике помоћу стандардне девијације X.

$$X^* = \frac{X - \bar{X}}{\sigma_X} \tag{6.3}$$

Подаци који су мањи од средње вредности имаће негативан, а они већи позитиван *z*-скор. Вредности променљивих које падају тачно на средњу вредност имаће z-скор нула. Вредности при овом начину стандардизације се најчешће креће у опсегу између -4 и 4 *(Larose, 2014)*. Овај метод је користан када максимум и минимум нису познати или када постоје изузеци који доминирају у мин-макс нормализацији *(Han et al., 2011).*

Дељење са максимумом је позната нормализациона функција коју су први предложили *Milligan & Cooper (1988)*, а касније *Gan et al. (2007)* за кластеровања података. Ова функција скалира све вредности са највећом вредношћу:

$$X^* = \frac{X}{\max\left(X\right)} \tag{6.4}$$

Модификована верзија ове функције додаје свим вредностима апсолутну вредност минимума, а онда скалира све највећом вредношћу у помереним подацима:

$$X^* = \frac{X + \left|\min(X)\right|}{\max\left(X + \left|\min(X)\right|\right)} \tag{6.5}$$

Код ове нормализационе функције долази до промене дистрибуције података. Такође, уколико у подацима постоје негативне вредности и оне ће се превести на јединични интервал. Са друге стране, то некада може бити лоше јер ће се може десити да две вредности, једна позитивна, а друга негативна, добију исту трансформисану вредност, па може доћи до губљења информација.

У овом истраживању коришћене су следеће нормализационе функције мин-макс, дељење са максимумом, и њена модификација.

6.5. Евалуације успешности класификације

При евалуацији успешности класификације важно је изабрати одговарајући метод за поделу података на тренирање и тестирање. Постоји више метода које се тиме баве *(Han et al., 2011; Koloseni, 2015).*

Holdout метода дели податке на случајан начин на два независна скупа за тренирање и тестирање. Тачност је веома зависна од састава података који се налазе у скупу за тренирање и тестирање, па процења тачност није увек довољно поуздана.

Случајно подузорковање је варијација *holdout* методе у којој се *holdout* метода понавља *k* пута, а процена укупне тачности се добија као просек тачности у свакој итерацији *(Han et al., 2011)*. Међутим, основна мана је што се скупови података за тестирање не формирају у складу са расподелом података. Да би се очувала расподела може се применити стратификација која ће прерасподелити податке тако да сваки подузорак буде добар представник целине (*Koloseni, 2015)*.

Унакрсна валидација је метода код које је основна идеја да се цео узорак искористи за валидацију. За разлику од претходних метода свака инстанца се користи исти број пута и за тренирање и за тестирање. Једна од најчешће коришћених је *k-fold* унакрсна валидација где се цео узорак дели на *k* приближно једнаких подузорака који су међусобно искључиви. Један подузорак се користи за тестирање, а преосталих *k-1* за тренирање, и то се се понавља *k* пута. Подслучајеви унакрсне валидације су *leave-one-out* метода и стратификована унакрсна валидација. Ради процене успешности класификације, у овом истраживању је коришћена унакрсна валидација.

За проблем класификације постоје неколико мера успешности. Избор одговарајуће мере највише зависи од скупа података односно од балансираности расподеле класа у узорку. Једна од најчешће коришћених мера је тачност класификације која представља проценат тачно класификованих инстанци у скупу података за тестирање.

Поред тога, матрица конфузије је користан алат за анализу колико добро класификатор препознаје инстанце различитих класа *(Han et al., 2011)*. У Табели 7 приказана је матрица конфузије за две класе:

Табела 7. Матрица конфузије за две класе

		Предвиђена	
		Класа 1	Класа 2
Стварна	Класа 1	C_{11} *(true positive)*	C_{12} *(false negative)*
	Класа 2	C_{21} *(false positive)*	C_{22} *(true negative)*

Елементи матрице конфузије имају следеће значење:

- C_{11} је број коректних предвиђања да је инстанца позитивна
- C_{22} је број коректних предвиђања да је инстанца негативна
- C_{21} је број погрешних предвиђања да је инстанца позитивна
- C_{12} је број погрешних предвиђања да је инстанца негативна

C_{21} се назива грешка типа 1, а C_{12} грешка типа 2. Грешка типа 1 је критична са аспекта доносиоца одлуке. Поред тачности и стопе грешке, из матрице конфузије се могу извести и друге мере као што су сензитивност, специфичност, итд.

Међутим, тачност није довољна мера за оцену успешности модела у случају небалансиране расподеле класа. Зато је потребно користити меру успешности која не зависи од балансираности узорка као што је област испод *ROC* криве.
У овом истраживању коришћени су балансирани скупови података па је као мера успешности изабрана матрица конфузије.

6.6. Карактеристике доброг класификатора

При избору или изградњи класификатора потребно је разматрати одговарајуће карактеристике класификатора који су најрелевантније са аспекта решавања одређеног проблема. Према *(Simpson, 1992; Koloseni, 2015)* пожељно је да добар класификатор поседује следеће особине:

- **Нелинеарна сепарабилност**. Потребно је да добар класификатор има способност да раздваја класе било ког облика и величине, укључујући нелинеарне.
- **Време за обучавање**. Добар класификатор не би требало да потроши превише времена за тренирање нелинеарних граница одлучивања.
- **Онлајн адаптација** је способност да класификатор учи нове класе и ажурира постојеће класе без уништавања старих информација. Офлајн адаптација захтева поновно тренирање целог система заједно са новим и старим информацијама сваки пут када се нова информација додаје у систем што захтева велику меморију и дуже време тренирања.
- **Преклапање класа.** Добар класификатор би требало да има способност да формира границу одлучивања која минимизира грешку класификације насталу услед преклапања класа.
- **Верификација и валидација.** Веома је важно да постоје адекватне мере за оцену перформанси било да је реч о графичким представама или о квантитатвним показатељима.
- **Параметри за подешавање.** Добар класификатор би требало да има што је мање могуће параметара за подешавање. Параметри би требало да се подесе на почетку, а не током тренинга. Поред тога важно је добро разумети ефекат параметара.
- **Непараметарска класификација.** Да би били поуздани, параметарски класификатори захтевају априори знање о функцији густине вероватноће за сваку класу. Међутим, априори знање није увек доступно. Због тога је важно да се код непараметарских класификатора дистрибуција података опише на начин који обезбеђује поуздану границу класа.
- **Меко и тврдо одлучивање**. Класификатор има способност да меко одлучује ако обезбеђује степен са којим одређени патерн одговара некој класи. У препознавању објеката могуће је да се систем тренира да препознаје класе примарних боја, а да користи меке процесе одлучивања да одреди степен до кога је свака од три боје присутна. Тврда одлука је када класификатор има способност да уочи да ли објекат припада некој класи или не.

Такође је идентификовано и више критерума за поређење резултата класификације. Према *(Han et al., 2011)*, методе класификације и предвиђања могу се оцењивати и поредити према следећим критеријумима:

- **Тачност** указује на способност класификатора да коректно предвиђа лабелу класе нових или претходно непознатих података. Тачност се може оцењивати помоћу једног или више скупова података за тестирање који су независни од тренинг скупа. Технике за оцену тачности као што је унакрсна валидација већ су претходно објашњене.
- **Брзина** подразумева да при генерисању и коришћењу класификатори не троше превише ресурса.
- **Робусност** је способност класификатора да коректно предвиђа ако су дати подаци са шумом или са недостајућим вредностима.
- **Скалабилност** се односи на способност да се ефикасно конструише класификатор у случају огромне количине података.
- **Интерпретабилност** се односи на ниво разумевања и могућности увида који класификатор омогућава. Интерпретабилност је субјективне природе и због тога је тежа за оцену.

Може се уочити да су неки од ових захтева супростављени, нпр. често се увођењем параметра повећава тачност али и време за обучавање. Са друге стране, неке од наведених особина неће бити релавантне за приступ класификацији који се у овом раду предлаже или излазе ван опсега ове дисертације, те се неће разматрати. Приступ класификацији заснованој на ИБА сличности евалуираће се са аспекта релевантних особина у оквиру експеримента у наставку рада.

7. ЛОГИЧКИ ПРИСТУП КЛАСИФИКАЦИЈИ

По угледу на друга истраживања која се баве увођењем вишевредносних мера поређења *(Luukka, 2005; Saastamoinen, 2006),* и у овој тези користи се класификација за евалуацију предложене параметризоване ИБА мере сличности и за њено поређење са другим логичким мерама сличности. Код *(Saastamoinen, 2006)* као један од основних мотива за коришћење класификације као оквира за тестирање наводи се фази природа проблема класификације, где питање није да ли објекат припада класи или не, већ у ком степену посматрани објекат припада некој класи. О фази природи класификације патерна и кластеровања говори *Zadeh (1977)* у једном од својих раних радова. У психологији *Tversky (1977)* наводи да се сличност користи као принцип организовања помоћу кога појединци класификују објекте, формирају концепте и заснивају генерализацију.

У овом поглављу, уводи се логичка класификација након чега се врши валидација параметарске ИБА мере сличности коришћењем класификације.

7.1. Логичка класификација заснована на ИБА

Прво се представља логички приступ класификацији заснован на ИБА (ЛК). Алгоритам логичке класификација се заснива на једноставним алгоритмима најближег прототипа и *k*-најближих суседа. Разлика је у томе што се код ЛК користи нова интерпретабилна ИБА мера сличности, оператор логичке агрегације као и различити начини поређења објеката при агрегацији. Будући да је веома експресиван алат, оператор логичке агрегације омогућава да се било која логичка функција користи за агрегацију сличности. Логичке функције могу бити задате и од стране експерта. Поред тога, овде постоје и два приступа поређењу објеката. Код првог приступа објекти се пореде по истим особинама коришћењем ИБА мере сличности, а затим се сличности агрегирају коришћењем оператора логичке агрегације. Код другог приступа, прво се формира логичка функција репрезентације објекта коришћењем логичке агрегације. То је логичка функција која најбоље описује неки објекат и могућу интеракцију између његових атрибута.

Може се тумачити и као изведена особина. Затим се врши поређење објеката по изведеној особини коришћењем ИБА мере сличности.

Због наведених својстава ЛК омогућава конструисање различитих класификатора који носе семантичку информацију о резултатима класификације. Чест случај у научној литератури је да се у први план ставља висока тачност класификације као једна од најпожељнијих особина добрих класификатора. Међутим, у многим применама једнако је важна и могућност интерпретације резултата класификације. За многе менаџере који користе алате за подршку одлучивању важно је да могу да разумеју резултат који су добили као излаз из система *(Kriegel& Schubert, 2006)*.

ЛК има добру математичку основу коју наслеђује од ЛА и ИБА мере сличности. Такође наслеђује једноставност од алгоритама за класификацију. Генерално, најближи прототип и *k*-најближих суседа су једноставни алгоритми код којих је учење засновано на инстанцама. За њих важи да се без потешкоћа примењују на различите проблеме јер не постоји модел који се обучава. У претходном поглављу 6 описане су особине алгоритама који се заснивају на лењом учењу.

7.1.1. Алгоритам логичке класификације

Основни задатак је да се класификује скуп података X у n различитих класа $C_1,\ldots,C_n$. Сваки објекат односно инстанца $x \in X$ је описана са m различитих атрибута $a_1,\ldots,a_m$.

Логичка класификација заснована на најближем прототипу (ЛКНП) се састоји из неколико корака. У првом кораку се врши *скалирање/нормализација скупа података за обучавање и тестирање на јединични* $[0,1]$ *интервал*. Ово је неопходан услов за примену ИБА мере сличности и логичке агрегације. Други корак се односи на *одређивање прототипа за сваку класу*. Вектори прототипова p_i за сваку класу $i=1,\ldots,n$ представљени су агрегираним атрибутима који се добијају логичком агрегацијом по атрибутима свих инстанци из скупа подака за тренирање димензије d :

$$p_i\left(a_1,\ldots,a_m\right)=\left(LA_{j=1}^{d}\left(a_{j1}\right),\ldots,LA_{j=1}^{d}\left(a_{jm}\right)\right),\quad i=1,\ldots,n \tag{7.1}$$

При формирању вектора прототипова најчешће се користи аритметичка средина по атрибутима. У овом моделу, аритметичка средина је само специјалан случај ЛА. Тиме се оставља простор за агрегацију атрибута неким другим ЛА оператором, или за генерисање више прототипова за једну класу коришћењем парцијалних агрегационих функција над атрибутима. У наредном кораку одређује се сличност између скупа података за тестирање и прототипова на следећи начин:

$$S_{IBA}^{LA_1}\left(x,y\right)=LA_{k=1}^{m}\left(1-x\left(a_k\right)-y\left(a_k\right)+2*\min\left(x\left(a_k\right),y\left(a_k\right)\right)\right),\quad \left(x,y\right)\in\left[0,1\right]^m \tag{7.2}$$

ЛА оператор може имати велики број реализација у зависности од података односно проблема, као и експертског знања. На пример, уколико у подацима постоји значајна корелација између одређених атрибута, један од начина да се моделује њихов утицај је да се за оператор генерализованог производа изабере минимум који се користи за атрибуте исте по природе. То је значајно приликом конструисања експертских логичких функција, што је показано у случају предвиђања банкротства у глави 8.

Претходни израз 7.2 се односи на поређење објеката по атрибутима, након чега се врши агрегација сличности. Међутим, у овом кораку одређивање сличности могуће је вршити и поређењем између логичких функција које представљају објекат на следећи начин:

$$S_{IBA}^{LA_2}\left(x,y\right)=1-LA_{k=1}^{m}x\left(a_k\right)-LA_{k=1}^{m}y\left(a_k\right)+2*\min\left(LA_{k=1}^{m}x\left(a_k\right),LA_{k=1}^{m}y\left(a_k\right)\right),\quad \left(x,y\right)\in\left[0,1\right]^m \tag{7.3}$$

У овом случају функција добијена логичком агрегацијом може се посматрати као изведени атрибут па се и овде објекти пореде по истој особини која је у ствари изведена особина. Могуће је дефинисати више различитих логичких функција репрезентације истог објекта у зависности од самих података и експерског знања. Са друге стране, јасно је да се при поређењу између прототипова и тест инстанци мора користити иста репрезентација.

У последњем кораку *додељује се класа инстанцама из скупу података за тестирање* на основу највеће сличности са прототипом. Формално тест инстанца x припада класи C_n ако има највећу сличност са прототипом који представља ту класу:

$$S_{IBA}^{LA}\left(x, p_n\right) = \max_{i=1,\dots,n} S_{IBA}^{LA}\left(x, p_i\right) \tag{7.4}$$

Друга група класификатора идејно полази од методе *k*-најближих суседа. Логичка класификација заснована на *k*-NN има сличне кораке са том разликом што се овде не одређује центроид за сваку класу већ се сличност рачуна између сваке две инстанце скупа за тренирање и тестирање помоћу једначина 7.2 и 7.3. Код овог алгоритма се некој тест инстанци додељује класа из скупа за обучавање која се најчешће појављује у околини *k*-најближих суседа те инстанце.

7.2. Валидација параметарских ИБА мера сличности на проблему класификације

Природно се поставља питање да ли увођење параметра може да унапреди перформансе класификације – тачност и друге мере успешности класификације. Основна идеја овог експеримента је да се на проблему класификације евалуира увођење параметра код ИБА мере сличности. Мотив, графичка интерпретација, као и математичка основа за увођење параметра, већ су изложене у претходном поглављу, а у овом поглављу идеја је да се емпиријски покаже оправданост параметризације.

7.2.1. Подаци

Подаци за тестирање преузети су са онлајн репозиторијума - *UCI Machine Learning Repository (Lichman, 2013)*. *UCI* репозиторијум је колекција база података које се користе за емпиријске анализе алгоритама машинског учења. Од свог настанка 1987. године студенти, предавачи и научници широм света користе овај репозиторијум као примарни извор скупова података за машинско учење.

За потребе овог експеримента изабрани су подаци који су коришћени у сличним истраживањима где се класификација заснивала на сличности нпр. (*Luukka & Leppalampi, 2006*). Коришћено је осам скупова података већином из домена медицине који се разликују према броју класа, атрибута и инстанци. Основне карактеристике сваког скупа података приказане су у Табели 8, а за сваку од њих дата су и детаљнија објашњења. Ови скупови података се могу сматрати базичним за тестирање перформанси класификатора.

Табела 8. Основне карактеристике тестираних скупова података

Подаци	Број класа	Број атрибута	Број инстанци
Iris	3	4	150
Wine	3	13	178
Dermatology - диференцијална дијагноза кожних болести	6	33	358
PIMA - дијагноза дијабетеса	2	8	768
BUPA - дијагноза поремећаја јетре	2	6	345
BCW1 - дијагноза тумора дојке	2	10	699
WDBC - дијагноза тумора дојке	2	31	569
Thyroid - дијагноза тироидне жлезде	3	5	215

Iris *(Fisher, 1936)* је једна од најпознатијих и најчешће коришћених база падатака у области препознавања патерна. Скуп података садржи 3 класе по 50 инстанци, а свака класа представља одређени тип биљке ириса. Једна класа је линеарно сепарабилна у односу на друге две, а последње две нису међусобно линеарно сепарабилне.

Wine *(Aeberhard et al., 1992)* је скуп података настао као резултат хемијске анализе вина која се гаје у истој области Италије али су преузета од три различита узгајивача. Анализа је одредила количине 13 различитих састојака који су пронађени у сваком од ова 3 типа вина. Број инстанци у класи 1 је 59, у класи 2 71, а у класи 3 је 48.

Скуп података **Dermatology** *(Güvenir et al., 1998)* се користи за дијагнозу кожних болести из групе планоцелуларног карцинома (*енг.* erythemato-squamous). Болести у овој групи су псоријаза (*енг.* psoriasis), себореични дерматитис (*енг.* seboreic dermatitis), лишајеви (*енг.* lichen planus), питиријазис росеа (*енг.* pityriasis rosea), хронични дерматитис (*енг.* cronic dermatitis) и пиларна црвена питиријаза (*енг.* pityriasis rubra pilaris). Обично је биопсија неопходна за дијагнозу али ове болести такође деле многе хистопатолошке карактеристике. Још један проблем код диференцијалне дијагнозе је што неко обољење може показивати карактеристике другог обољења у почетној фази и може имати карактеристична својства у каснијим фазама.

PIMA *(Smith et al., 1988)* је скуп података на основу кога се испитује присуство дијабетеса међу Пима Индијанцима, односно Пима женама које су старије од 21. године. Класе 1 садржи 500 инстанци негативних на дијабетес, док су инстанце класе 2 позитивне на дијабетес и има их 268.

BUPA су подаци донирани *UCI* репозиторијуму *(Lichman, 2013)* од стране *R. S. Forsyth*-а. Део су медицинског истраживање где је циљ да се предвиди да ли мушки пацијент има поремећај јетре. Првих 5 атрибута су тестови крви за које се сматра да указују на поремећај јетре који је настао услед прекомерног конзумирања алкохола. Шести атрибут је количина попијених алкохолних напитака дневно у полу-пинта еквиваленту. Класа 1 садржи 145, а класа 2 200 инстанци.

BCW1 је скуп података настао из медицинског истраживања *(Wolberg & Mangasarian, 1990)*, где је покушано да се само на основу биопсије танком иглом (*енг.* fine needle aspiration) прецизно установи дијагноза ткива дојке. Визуелном проценом узорка биопсије идентификовано је 9 релевантних атрибута за дијагнозу. **WDBC** је настао у оквиру наредног истраживања *(Street et al., 1993)*, где су карактеристике израчунате на основу дигиталне слике ткива дојке. Бенигна класа садржи 357, а малигна класа 212 инстанци.

Скуп података **Thyroid** *(Coomans et al., 1983)* садржи резултате 5 лабораторијских тестова. На основу њих се предвиђа да ли пацијент има нормалну, хипо или хипер функцију штитне жлезде. Дијагноза се успоставља на основу комплетног медицинског картона као што је анамнеза, скенер, итд.

7.2.2. Експеримент - класификација са различитим логичким мерама сличности

Тестирање је замишљено тако да се користи класификација заснована на прототипу у оквиру које се мењају различите логичке мере сличности. *Saastamoinen (2006)* је у свом истраживању емпиријски показао да мере поређења засноване на вишевредносној логици могу дати једнако добру тачност класификације као и метричке мере сличности на тестираним медицинским подацима. Другим речима, показало се да у практичним применама испуњење метричких аксиома има мало или нимало утицаја на резултате. Због тога се у оквиру овог истраживања неће вршити поређење са метричким мерама сличности, већ само са логички заснованим мерама. Прво се представљају резултати поређења са минимум, производ и Лукашијевич мерама сличности (једначине 2.1, 2.2 и 2.3) који представљају подслучајеве код параметарске ИБА мере сличности (једначина 5.3). Поред тога се за поређење користи и мера сличности заснована на генерализованој Лукашијевич структури и генерализованом просеку (ГЛСГП) која је једна од најчешће коришћених и најуспешнијих логичких мера поређења (*Luukka & Leppalampi, 2006*). Ова мера је двопараметарска (зависи од параметара *m* и *p*) и има следећи облик:

$$S_{GLSGP}^{p,m}\left(x_i, y_i\right) = \left(\sum_{i=1}^{n} \left(1 - \left|x_i^{\,p} - y_i^{\,p}\right|\right)^{\frac{m}{p}} \right)^{\frac{1}{m}} \tag{7.5}$$

Да би посматрана логичка мера сличности дошла до изражаја, а не утицај оператора агрегације, примењена је аритметичка средина за агрегацију сличности. Већ је познато да је аритметичка средина специјалан случај ЛА *(Radojević, 2008)*. Код ГЛСГП средња вредност за оператор агрегације се добија за вредност параметра *m*=1 у генерализованој средњој вредности. Тиме су обезбеђени једнаки услови за поређење свих логичких мера сличности. За тестирање ГЛСГП коришћен је програмски код из рада *(Luukka & Leppalampi, 2006)*. У Матлаб програмском окружењу додатно је имплементиран програм за тестирање непараметарских/параметарских ИБА мера сличности у комбинацији са различитим операторима логичке агрегације и за различите алгоритме класификације.

За нормализацију података на јединични $[0,1]$ интервал користе се три нормализационе функције описане у претходном поглављу код препроцесирања података (једначине 6.2, 6.4 и 6.5). Нормализације мин-макс (N_1) (једначина 6.2) и дељење са максимумом (N_2) (једначина 6.4) су стандардне $[0,1]$ нормализације које се често користе у литератури. Обе су линеарне трансформације односно чувају однос који постоји у оригиналним подацима *(Pyle, 1999)*. Z-скор, односно стандардизација нула-средња вредност, није била одговарајућа због опсега. Последња нормализација која је изабрана је модификована нормализација дељење са максимумом (N_3) (једначина 6.5) која утиче на расподелу података у смислу да помера нормализоване вредности ка горњем делу јединичног интервала.

Унакрсна валидација је коришћена за евалуацију успешности класификације. Узорак се делио на случајан начин на два дела за тренирање и тестирање у односу 50%-50% и 70%-30%. Овај поступак се понављао 100 пута за сваку комбинацију мере сличности, вредности параметра и нормализационе функције. Параметри m и p код тестирања генерализоване Лукашијевич структуре са генерализованим просеком узимали су вредности из предложених интервала у истраживању *(Luukka & Leppalampi, 2006)*, где су дате најбоље области параметара за сваки медицински скуп података. За параметарску ИБА меру сличности тестирање је вршено за вредности параметара α и β на интервалу $[0,1]$ како је претходно и дефинисано у поглављу 5.

7.2.3. Резултати и анализа

У овом делу рада су представљени резултати 5 експеримената где се ИБА мере сличности тестирају на проблему класификације.

Е1: Први експеримент је имао за циљ да упореди резултате класификације засноване на логичким мерама сличности: минимум, производ и Лукашијевич би-импликацију са ИБА еквиваленцијом – непараметарском ИБА мером сличности.

У експерименту се показало да се приближна тачност класификације добија за поделу узорка на 50%-50% и 70%-30%. Затим, од три врсте нормализационих функција које су коришћене, показало се да код медицинских података последња N_3 даје нешто боље резултате у односу на N_1 и N_2. Стога, у Табели 9 приказани су резултате класификације за нормализациону функцију N_3 са поделом узорка на 70%-30% за 100 понављања. Подаци о просечној тачности и просечној варијанси (Вар.) за остале две нормализационе функције приложени су у Прилогу А.

Табела 9. Резултати класификације за непараметарске логичке мере за нормализацију N_3 и поделу 70%-30%

	S_M		S_P		S_L		S_{IBA}	
	Тачност	Вар.	Тачност	Вар.	Тачност	Вар.	Тачност	Вар.
Iris	0,5282	0,0024	0,7353	0,0034	0,9569	0,0007	0,9610	0,0006
Wine	0,3273	0,0000	0,3775	0,0020	0,9456	0,0006	0,9429	0,0006
Dermatology	0,1855	0,0002	0,1891	0,0005	0,9582	0,0003	0,9550	0,0004
PIMA	0,3523	0,0000	0,4342	0,0005	0,7336	0,0007	0,7277	0,0008
BUPA	0,4235	0,0000	0,4247	0,0005	0,5483	0,0019	0,5463	0,0020
BCW1	0,9110	0,0002	0,9271	0,0003	0,9428	0,0002	0,9417	0,0002
WDBC	0,5738	0,0007	0,8163	0,0006	0,9297	0,0004	0,9358	0,0003
Thyroid	0,7214	0,0003	0,7423	0,0008	0,9438	0,0009	0,9417	0,0009

Очекивано, тачност класификације се повећава од минимум до Лукашијевич би-импликације односно ИБА еквиваленције. Незнатна разлика која постоји између ИБА мере сличности и Лукашијевич би-импликације је настала искључиво као резултат случајне поделе узорка, пошто су ове мере математички еквивалентне. Такође, посматрајући вредности варијансе очекивано мања варијанса се добија за ИБА еквиваленцију и Лукашијевич би-импликацију, у односу на минимум и производ би-импликацију.

E2: У другом експерименту се посматрала класификација са ИБА параметарском мером сличности са једним параметром $\alpha \in [0,1]$ (једначина 5.3). За поређење коришћена је мера заснована на генерализованој Лукашијевич структури са генерализованим просеком (једначина 7.5) као пример успешне параметарске логичке мере. Да би се обезбедило што равноправније поређење, за оператор логичке агрегације изабрана је аритметичка средина, а код ГЛСГП параметар $m=1$.

Овај експеримент је имао за циљ да тестира да ли ће параметарска ИБА мера сличности унапредити перформансе класификације ако се параметри оцене из података. Критеријумска функција је тачност класификације, а задатак је наћи вредност параметра мере сличности којој одговара највећа просечна тачност класификације. За параметарску ИБА меру сличности вредност параметра је увек на јединичном интервалу $\alpha \in [0,1]$, док се параметар p код ГЛСГП бира из одговарајућег опсега за сваки скуп података посебно. Секвенцијално се за сваку вредност параметра из посматраног опсега израчунава просечна тачност класификације за 100 понављања где се сваки пут узорак дели на случајан начин. На крају се бира она вредност параметра којој одговара највећа просечна тачност класификације.

Табела 10. Резултати класификације за параметарске логичке мере за нормализацију N_3 и поделу 70%-30%

	S_{IBA}		S_{IBA}^{α}			$S_{GLSGP}^{p,1}$		
	Тачност	Вар.	Тачност	Вар.	α	Тачност	Вар.	p
Iris	0,9610	0,0006	0,9610	0,0006	1,00	0,9570	0,0005	0,50
Wine	0,9429	0,0006	0,9438	0,0007	0,95	0,9469	0,0007	0,70
Dermatology	0,9550	0,0004	**0,9775**	0,0001	0,40	0,9634	0,0003	1,30
PIMA	0,7277	0,0008	0,7321	0,0005	0,92	0,7433	0,0004	2,80
BUPA	0,5463	0,0020	**0,5776**	0,0020	0,04	0,5476	0,0017	9,60
BCW1	0,9417	0,0002	**0,9737**	0,0001	0,50	0,9549	0,0002	3,35
WDBC	0,9358	0,0003	0,9387	0,0003	0,80	0,9380	0,0003	2,25
Thyroid	0,9417	0,0009	**0,9615**	0,0004	0,60	0,9438	0,0009	1,00

И у другом експерименту нормализациона функција N_3 дала је најбоље резултате за све посматране мере, а различите поделе података 50%-50% и 70%-30% за тренинг и тестирање нису направиле значајну разлику.

У Табели 10 види се да је просечна тачност класификације за параметарску ИБА меру сличности иста или боља у односу на обичну непараметарску. Боља тачност је добијена за половину скупова података (Dermatology, BUPA, BCW1 и Thyroid).

Последње две колоне Табеле 10 се односе на поређење две параметарске мере. Овде се показало да је параметарска ИБА мера сличности за већину скупова података једнако добра или нешто боља од ГЛСГП. Параметарска ИБА мера је боља за четири скупа података (Dermatology, BUPA, BCW1 и Thyroid), док ГЛСГП има незнатно бољу тачност само за PIMA. Као што је раније показано, значај параметарске ИБА мере сличности се свакако огледа у њеној јасној интерпретацији. Такође, параметарска ИБА еквиваленција је математички једноставнија и захтева мање времена за израчунавање у односу ГЛСГП. У Табели 13 дати су резултати поређења параметарских мера према времену израчунавања.

Е3: За трећи експеримент уводи се двопараметарска ИБА мера сличности са параметрима који су линеарна конвексна комбинација уз "имање" и "немање" особине:

$$S_{IBA}^{\alpha,1-\alpha}(x,y)=\left[1-x-y+\min(x,y)\right]\cdot(1-\alpha)+\min(x,y)\cdot\alpha \tag{7.6}$$

У Табели 11 су упоређени резултати класификације добијени за ову меру са резултатима класификације за једнопараметарску ИБА меру сличности.

На медицинским подацима се показало да ИБА мера сличности са линеарно конвексном комбинацијом параметра α даје приближно исте резултате као и са једним параметром. Другим речима, ова линеарна веза не доприноси тачности класификације у случају тестираних медицинских података.

Табела 11. Резултати класификације за параметарске логичке мере за нормализацију N_3, поделу 70%-30% и α, 1-α

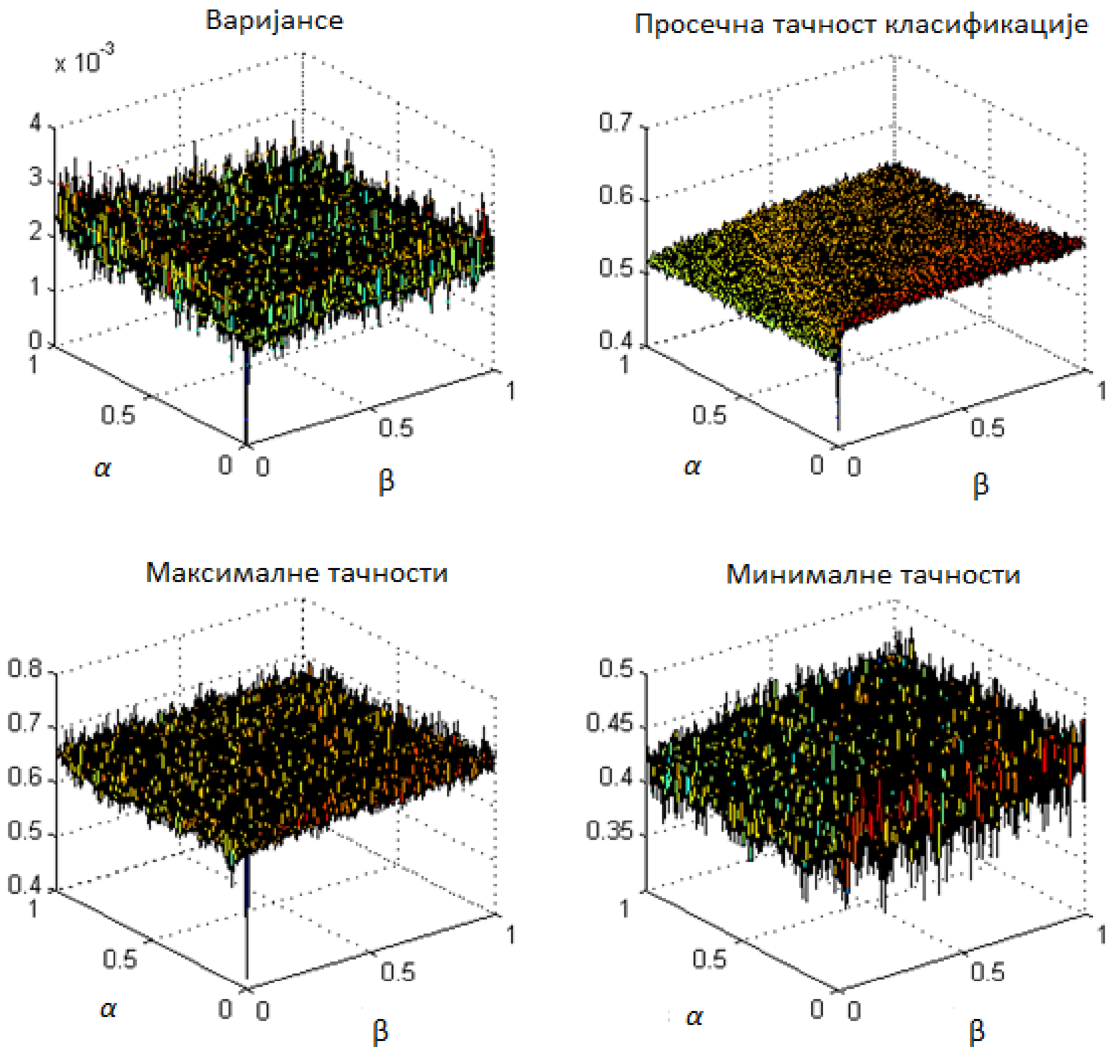
	S_{IBA}^{α}			$S_{IBA}^{\alpha,1-\alpha}$		
	Тачност	Вар.	α	Тачност	Вар.	α
Iris	0,9610	0,0006	1,00	0,9590	0,0006	0,50
Wine	0,9438	0,0007	0,95	0,9411	0,0008	0,50
Dermatology	0,9775	0,0001	0,40	0,9764	0,0002	0,60
PIMA	0,7321	0,0005	0,92	0,7409	0,0008	0,45
BUPA	0,5776	0,0020	0,04	0,5754	0,0018	1,00
BCW1	0,9737	0,0001	0,50	0,9727	0,0001	0,75
WDBC	0,9387	0,0003	0,80	0,9353	0,0003	0,55
Thyroid	0,9615	0,0004	0,60	0,9665	0,0004	0,65

Е4: У четвртом експерименту се код параметарске ИБА мере сличности уводе два параметра која су независна. Ови резултати су у Табели 12 упоређени са резултатима класификације са двопараметарском ГЛСГП.

Табела 12. Резултати класификације за параметарске логичке мере за нормализацију N_3, поделу 70%-30% и α, β

	$S_{GLSGP}^{p,m}$			$S_{IBA}^{\alpha,\beta}$		
	Тачност	Вар.	$(p;m)$	Тачност	Вар.	$(\alpha;\beta)$
Iris	0,9687	0,0005	(1,05; 0,60)	0,9662	0,0005	(0,25; 0,26)
Wine	0,9556	0,0005	(0,40; 0,40)	0,9513	0,0006	(0,89; 0,97)
Dermatology	0,9620	0,0002	(1,15; 1,00)	0,9795	0,0002	(0,64; 0,99)
PIMA	0,7587	0,0005	(2,30; 2,15)	0,7475	0,0005	(0,17; 0,14)
BUPA	0,5675	0,0016	(9,65; 1,30)	**0,5918**	0,0023	(0,00; 0,54)
BCW1	0,9734	0,0001	(0,15; 0,05)	0,9767	0,0001	(0,31; 0,88)
WDBC	0,9547	0,0002	(3,00; 0,01)	0,9415	0,0003	(0,35; 0,43)
Thyroid	0,9451	0,0009	(1,00; 1,00)	**0,9669**	0,0003	(0,18; 0,32)

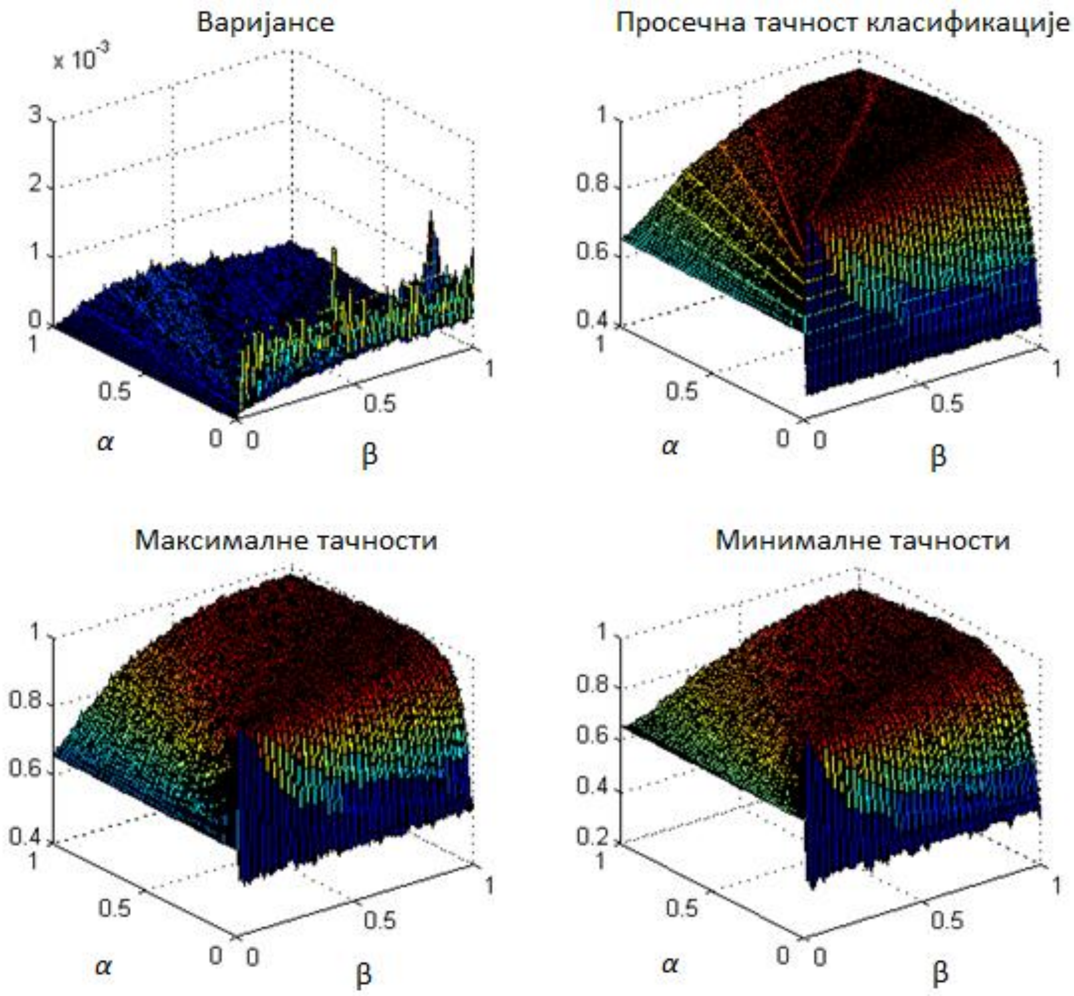
Обе мере са два параметра дају приближне резултате за тачност класификације, али је ИБА еквиваленција са два параметра нешто боља код BUPA и Thyroid. Поред тога, посматрана је и зависност тачности класификације од вредности параметара за ИБА меру сличности и за ГЛСГП. У наставку су приказане зависности тачности класификације на изабраним скуповима података BUPA, BCW1 и Dermatology за параметарску ИБА меру сличности на основу којих се могу донети јасни закључци. Остали графички прикази зависности за ИБА меру сличности и за ГЛСГП су приложени у Прилозима Б и В.



Слика 15. Зависност тачности од параметара код ИБА мере сличности – BUPA

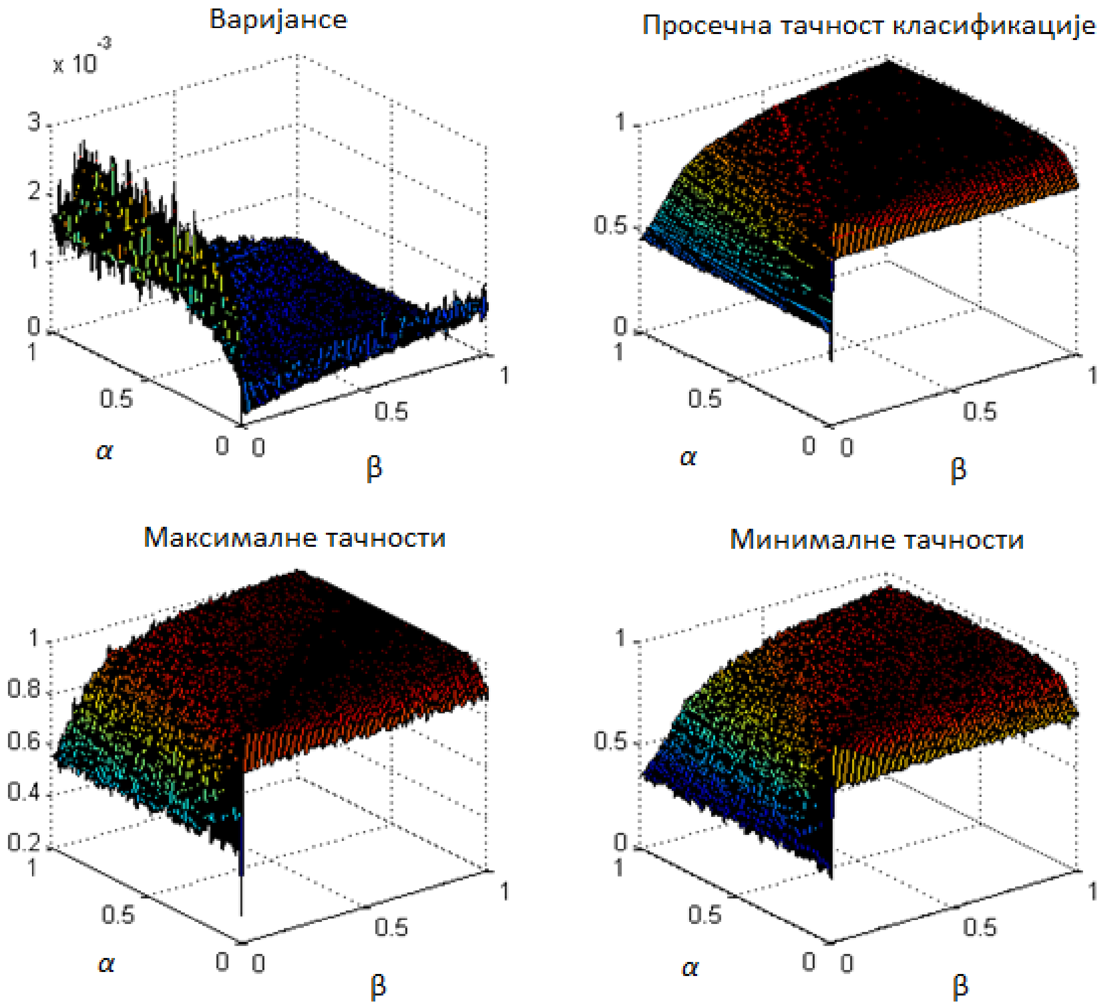
На Сликама 15, 16 и 17 приказане су површи које показују како промена параметара утиче на варијансу, најмање, највеће и просечне тачности класификације.

Иако је увођењем параметара код ИБА мере сличности код BUPA дошло до благог повећања тачности, на Слици 15 се јасно види да је промена вредности параметара имала врло мало утицаја на резултате класификације.



Слика 16. Зависност тачности од параметара код ИБА мере сличности – BCW1

На Сликама 16 и 17 приказане су површи зависности тачности класификације за BCW1 и Dermatology. На Слици 16 се види да вредности оба параметра значајно утичу на резултат класификације. Са друге стране, код Dermatology параметар β који се односи на имање особине има већи утицај (Слика 17). Стога, може се закључити да је узимање пуне вредности за имање особине обавезно при проучавању овог скупа података.



Слика 17. Зависност тачности од параметара код ИБА мере сличности – Dermatology

Када се резултати ИБА мере са два параметра пореде са једнопараметарском, може се закључити да на овим подацима увођење додатног параметра није значајно допринело тачности класификацији јер ИБА са једним параметром има незнатно лошије резултате од двопараметарске мере.

Из једначина 5.12 и 7.5 се јасно види да је параметарска ИБА еквиваленција математички једноставнија од ГЛСГП. Због тога се за обе мере тестирало време које је потребно за извршавање алгоритма за оцену параметара из података. Као скуп података за тестирање узет је Iris. Поред тога, потребно је напоменути да се за параметарске ИБА мере према дефиницији увек претражује јединични интервал у случају оба параметра, док је код ГЛСГП мере опсег вредности доста шири и зависи од скупа података. Да би се времена израчунавања поредила за једнак број итерација овде се за све параметре, било за ИБА или ГЛСГП мере, узима 100 дискретних вредности из јединичног интервала. Овакво ограничење не утиче на тачност класификације са ИБА мером, али је јасно да може негативно утицати на тачност класификације ГЛСГП мере (ова мера не мора да даје највећу тачност за вредности параметра са јединичног интервала, видети Табеле 11 и 12). За сваку вредност параметра S_{IBA}^{α} и $S_{GLSGP}^{p,1}$ или за сваку комбинацију два параметра код $S_{IBA}^{\alpha,\beta}$ и $S_{GLSGP}^{p,m}$ тренирање и тестирање се понављало 100 пута.

Табела 13. Поређење параметарских мера према времену извршавања

	S_{IBA}^{α}	$S_{GLSGP}^{p,1}$	$S_{IBA}^{\alpha,\beta}$	$S_{GLSGP}^{p,m}$
Време извршавања (s)	11,07	13,85	1126,85 (18,78 мин.)	1488,07 (24,80 мин.)

Као што је и очекивано, резултати у Табели 13 потврђују да је потребно 20% мање времена за оцену једнопараметраске ИБА мере у односу на ГЛСГП меру са једним параметром. Такође, потребно је 25% мање времена за оцену два параметра код ИБА мере у односу на двопараметарску ГЛСГП меру. Очигледно је да се ефикасност може унапредити за обе мере коришћењем одговарајућих оптимизатора, што ће бити предмет будућег истраживања.

Е5: Оно што је такође било занимљиво за тестирање и анализу је случај када се вредности параметра ИБА мере бирају из опсега од -1 до 1. Овде се графичка интерпретација мере може потенцијално проширити. У случају негативне вредности параметра, "немање" особине би променило смер и графички би се ширило ка делу који се односи на различитост. Међутим у том случају губе се својства мере сличности (нпр. услов о ненегативности), али се такође напушта ИБА као $[0,1]$ реализација Булове алгебре. Потенцијална решења су разматрана у поглављу 5.3. Проширење опсега није донело побољшање у односу на ИБА меру сличности са параметром од 0 до 1. Узевши у обзир да је ово истраживање у повоју, резултати класификације су приложени у Прилогу Г. Потребно је додатно истражити који су то карактеристични скупови података код којих би увођење параметра од -1 до 1 имало одговарајући утицај.

7.3. Закључак експеримента

Основни циљ је био да се потврди практична вредност дефинисаних параметарских ИБА мера сличности. Тестирање је вршено на примеру познатих медицинским подака за класификацију.

Спроведено је пет експеримената где је вршена логичка класификација заснована на најближем центроиду. Прво је дато поређење логичких мера сличности минимум, производ и Лукашијевич фази би-импликације са ИБА еквиваленцијом, где је очекивано потврђено да су последње две мере боље за мерење сличности од мера које се заснивају на минимум и производ би-импликацији.

Затим је вршено поређење параметарске ИБА мере са једном од најуспешнијих параметарских мера поређења – генерализованом Лукашијевич структуром са генерализованим просеком. Параметарске мере су разматране са аспекта тачности класификације, варијансе, времена израчунавања, утицаја параметара. Показало се да је просечна тачност класификације за параметарску ИБА меру сличности иста или боља у односу на обичну непараметарску. Може се закључити да увођење параметра доприноси тачности класификације у случају тестираних медицинских података. Другим речима, део израза у ИБА мери сличности који се односи на немање особине може се додатно прилагодити проблему.

При поређењу мера са једним параметром, показало се да је параметарска ИБА мера сличности за већину скупова података једнако добра или нешто боља од ГЛСГП. Међутим, обе мере са два параметра дају приближно исте резултате за тачност класификације. При анализи графичког приказа зависности резултата класификације од вредности параметара, могло се јасно уочити колико сваки од параметара ИБА мере сличности доприноси тачности за сваки посматрани скуп података. Такође, показано је да услед математичке једноставности ИБА мера захтева мање времена за израчунавање него ГЛСГП при оцени параметара.

Коначно, емпиријска анализа је потврдила практичну вредност параметарске ИБА мере, а посебно ИБА мере сличности са једним параметром због остварене тачности класификације, рачунске једноставности и јасне интерпретације.

8. ЛОГИЧКА КЛАСИФИКАЦИЈА ЗА ПРЕДВИЂАЊЕ БАНКРОТСТВА ПРЕДУЗЕЋА

Проблем предвиђања банкрота је један од најчешћих примера класификације у домену финансија. Предвиђање банкротства предузећа је сложен проблем. Постоји неколико отворених питања која се односе на временски период предвиђања, избор финансијских показатеља који се користе при анализи, веродостојност финансијских извештаја на којима се заснивају показатељи, анализу додатних нефинансијских фактора који могу бити значајни за предвиђање, итд. За потребе експеримента, у наставку је објашњен Алтманов *Z*-скор модел.

8.1. Алтманов *Z*-скор модел

Један од најчешћих модела за предвиђања банкрота је Алтманов *Z*-скор модел *(Altman, 1968)* који се заснива на линеарној дискриминантој анализи. Иако је Алтманов *Z*-скор једноставан модел показао се као веома ефикасан и већ неколико деценија је у употреби. У свом истраживању *Altman (1968)* је анализирао 22 финансијска показатеља и дискриминационом анализом одабрао 5 финансијских рација. Користио је податке 66 предузећа које је поделио на две групе: активна и банкротирала предузећа. Први Алтманов *Z*-скор модел се односи на јавне компаније и представља линеарну комбинацију следећих 5 финансијских показатеља:

$$Z = 1,2X_1 + 1,4X_2 + 3,3X_3 + 0,6X_4 + 0,999X_5 \tag{8.1}$$

где су:

X_1 = Нето обртна средства / Укупна актива

X_2 = Нераспоређена добит / Укупна актива

X_3 = Добит пре исплате камате и пореза / Укупна актива

X_4 = Тржишна вредност сопственог капитала / Књиговодствена вредност укупних обавеза

X_5 = Приход од продаје / Укупна актива

Алтманов *Z*-скор нижи од 1,81 указује на банкрот предузећа, а преко 2,99 да је мала вероватноћа банкрота. Опсег између 1,81 и 2,99 представља сиву зону.

У каснијим радовима *Altman (2000)* је представио нове верзије *Z*-скор модела. У случају приватних производних предузећа Z' модел је:

$$Z' = 0,717X_1 + 0,847X_2 + 3,107X_3 + 0,42X_4 + 0,998X_5 \tag{8.2}$$

Линеарни Алтманови *Z*-скор модели су коришћени као основа за логичке класификаторе који се конструишу у експерименту.

8.2. Подаци

У експерименту је коришћена база података компаније „*CUBE Risk Management Solutions*" која послује у сфери пружања пословних информација. База података садржи финансијске податке о 1020 средњих предузећа у Србији. Компаније су одабране у складу са чланом 6. Закона о рачуноводству („Сл. гласник РС", број 62/13) који се бави класификацијом легалних ентитета у Србији. Правна лица се разврставају у микро, мала, средња и велика правна лица, у зависности од броја запослених, пословног прихода и просечне вредности пословне имовине утврђених на дан састављања редовног годишњег финансијског извештаја у пословној години. У средња правна лица разврставају се она правна лица која задовољавају два од следећих критеријума:

- Просечан број запослених од 50 до 250;
- Пословни приходи од 8.800.000 до 35.000.000 ЕУР у динарској противвредности;
- Просечна вредност пословне имовине (израчуната као аритметичка средина вредности на почетку и на крају пословне године) од 4.400.000 до 17.500.00 ЕУР у динарској противвредности.

Финансијску базу чине подаци из годишњих финансијских извештаја биланса стања и биланса успеха за 2014. и 2015. годину. Од свих средњих предузећа у

Србији, после пречишћавања у бази су остали подаци за 1020 предузећа. Од тога 984 компаније су активне, а 36 су банкротирале. У Србији се под стечајем предузећа подразумева банкротство или реорганизација. У овом истраживању се анализирају предузећа која су банкротирала у складу са овом законском терминологијом. Иста база података је коришћена и за полазна истраживања логичког класификатора заснованог на ИБА *(Poledica et al., 2016).* За потребе овог истраживања коришћен је балансиран узорак од 72 компаније ради што равноправнијег поређења модела. За сваку компанију израчунато је 5 финансијских рација која се користе у Алтмановом *Z* односно *Z'* скору. Финансијска рација за 2014. годину су коришћена као улази за различите логичке класификаторе. Излаз из модела је предвиђени статус компаније у 2015. години: активан или банкрот (у стечају).

Да би се обезбедило адекватно поређење модела у експерименту било је важно узети у обзир и разне карактеристике података који могу имати утицај на перформансе класификације. У случају Алтмановог *Z* скора ови фактори могу бити година банкрота, величина компаније, старост компаније, индустрија и земља порекла *(Altman et al., 2016).* За што тачнију класификацију свог модела *Altman (2016)* предлаже коришћење података што ближе садашњем тренутку. Такође, значајан утицај на перформансе модела има и величина компаније. Да би се елиминисао овај ефекат при анализи користе се само средња предузећа. Поред тога *Altman (2016)* примећује да старост компаније може утицати на ризик од банкрота - веома младе компаније типично показују веома висок ризик стечаја. Међутим, овде се у анализи користе средња предузећа која углавном послују неко време и сматрају се релативно стабилним. Још један фактор од утицаја је земља порекла. С тим у вези, постоје различити економски и културни фактори, финансијска тржишта, прописи и рачуноводствена пракса који сви могу да утичу на финансијско понашање компанија. Алтманов *Z* скор није прилагођен српској економији.

У експерименту за предвиђање банкрота коришћен је Алтманов *Z* скор и *Z'* скор као бенчмарк модели при поређењу са осталим логичким класификаторима. Идеја је да се користе иста рација као и у једноставном, линеарном Алтмановом моделу да би се тестирало да ли је могуће логичким класификаторима односно моделовањем логичке интеракције унапредити перформансе класификације.

Такође, база података садржи негативне вредности атрибута, што није био случај код медицинских података. Због тога се користи одговарајућа функција нормализације N_3 (једначина 6.5) која има способност да преводе и негативне вредности на јединични интервал. Подела података на скуп за тренирање и тестирање 50%-50% и 70%-30% није имала значајан утицај на резултате класификације.

Различити логички класификатори који су конструисани у овом истраживању детаљно су описани у наставку.

8.3. Експеримент и резултати

Основна идеја овог дела истраживања је да се тестира да ли се конструисањем класификатора са експертским логичким функцијама могу унапредити перформансе класификације. У експериментима ЛК1 и ЛК2 користе се класификатори засновани на најближем центроиду и k-најближих суседа са непараметарском ИБА мером сличности и различитим операторима логичке агрегације. При конструисању класификатора примењена су и два начина поређења објеката у класификацији - према атрибутима објеката појединачно или у односу на агрегирану репрезентацију објеката (изведени атрибут) што је описано у поглављу 7.1 Логичка класификација заснована на ИБА. У оквиру експеримента ЛК3 приказују се резултати класификације за параметарске мере сличности за финансијски скуп података.

Добијени резултати класификације упоређени су са резултатима традиционалних метода за предвиђања банкрота (Алтманов Z и Z'-скор модел), приказаним у Табелама 14 и 15.

Табела 14. Резултати класификације за Алтман *Z* и *Z’*-скор

	Алтман *Z*-скор		Алтман *Z’*-скор	
	Активне	Банкротиране	Активне	Банкротиране
Активне	33 (45,83%)	8 (11,11%)	35 (48,61%)	9 (12,50%)
Банкротиране	3 (4,17%)	28 (38,89%)	1 (1,39%)	27 (37,50%)
Укупна тачност	0,8472		0,8611	

Табела 15. Резултати класификације за Алтман *Z* и *Z’*-скор укључујући и сиву зону

	Алтман *Z*-скор			Алтман *Z’*-скор		
	Активне	Сива зона	Банкр.	Активне	Сива зона	Банкр.
Активне	27 (37,5%)	6 (8,3%)	3 (4,2%)	20 (27,8%)	15 (20,8%)	1 (1,4%)
Банкротиране	6 (8,3%)	2 (2,8%)	28 (38,9%)	6 (8,3%)	3 (4,2%)	27 (37,5%)
Сумарно	33 (45,8%)	8 (11,1%)	31 (43,1%)	26 (36,1%)	18 (25,0%)	28 (38,9%)

Алтманови *Z* и *Z’* модели су једноставне параметарске једначине па је било могуће израчунати вредности за цео узорак од 72 предузећа. Такође, постоје дефинисане границе за три групе предузећа: активна, сива зона и банкрот. С обзиром да у доступним реалним подацима не постоји дефинисан статус предузећа „сива зона“, оним предузећима у узорку која према Алтману припадају сивој зони је додељено да су активна. Када се разматра сива зона, у Табели 15. види се да је значајно већи број предузећа из сиве зоне стварно и остао у статусу активан. Ово је битно имати у виду због каснијег поређења резултата класификације.

8.3.1. Резултати експеримента ЛК1

ЛК1.1 представља класификатор заснован на најближем центроиду који се рачуна као аритметичка средина за сваку класу. Поређење тест скупа са центроидом сваке класе врши се између атрибута објеката. Затим се врши агрегација сличности да би се добила укупна сличности.

ЛК1.2 Поред тога вршено је поређење и на основу агрегиране репрезентације објекта. Прво се добије изведена особина за сваки објекат па се после израчунавају сличности између изведених особина скупа за тестирање и центроида. Добијени резултати за ЛК1.1 и ЛК1.2 су приказани у Табели 16.

Табела 16. Резултати класификације за ЛК1.1 и ЛК1.2 за поделу 50%-50%

Нормализација	ЛК1.1		ЛК1.2	
	Тачност	Вар.	Тачност	Вар.
N_3	0,8231	0,0025	0,8167	0,0024

Узорак је балансиран, а коришћена је подела узорка 50%-50%. Оба наведена класификатора ЛК1.1 и ЛК1.2 нису дала боље резултате у односу на Алтманове моделе (видети Табеле 14 и 15). Најбољи остварени резултати за тачност класификације су приближно 82% што је лошије од *Z*-скор и *Z'*-скор модела.

8.3.2. Резултати експеримента ЛК2

У експерименту ЛК2 користе се класификатори засновани на *k*-најближих суседа са непараметарском ИБА мером сличности и различитим операторима логичке агрегације. Поред стандарног оператора аритметичке средине (**ЛК2.1**), идеја је да се у овом примеру моделује логичка интеракција између променљивих помоћу експертских логичких функција (ЛК2.2 и ЛК2.3).

ЛК2.2 има за циљ да укључи интеракцију између променљивих коришћењем алгебарског производа као оператора генерализованог производа између свих променљивих. Следећи генерализовани Булов полином

$$S_{IBA}^{\otimes}(x,y)=S_{IBA}\left(x(a_1),y(a_1)\right)\otimes S_{IBA}\left(x(a_2),y(a_2)\right)\otimes\ldots\otimes S_{IBA}\left(x(a_m),y(a_m)\right),\quad (x,y)\in[0,1]^m \quad (8.3)$$

у случају 5 финансијских рација као атрибута и заменом $\otimes := *$ постаје:

$$S_{IBA}^{prod}(x,y)=\prod_{k=1}^{5}\left(1-x(a_k)-y(a_k)+2*\min\left(x(a_k),y(a_k)\right)\right),\quad (x,y)\in[0,1]^5 \quad (8.4)$$

ЛК2.3 је сличан као претходни класификатор, са разликом што узима у обзир природу променљивих коришћењем оператора минимум $\otimes := \min$ између одговарајућих променљивих:

$$\begin{aligned} S_{IBA}^{\min,prod}(x,y) = \min\left[S_{IBA}\left(x(a_2),y(a_2)\right),S_{IBA}\left(x(a_3),y(a_3)\right)\right]\cdot \\ \cdot\min\left[S_{IBA}\left(x(a_2),y(a_2)\right),S_{IBA}\left(x(a_4),y(a_4)\right)\right]\cdot S_{IBA}\left(x(a_1),y(a_1)\right)\cdot S_{IBA}\left(x(a_5),y(a_5)\right) \end{aligned} \quad (8.5)$$

Овде се оператор минимум примењује као оператор при агрегацији атрибута a_2 и a_3, и a_2 и a_4 јер се показало да су ове променљиве међусобно високо корелисане што указује на сличну природу. Генерално, у анализи података опште је познато да постојање високо корелисаних променљивих у подацима може неоправдано да појача утицај неких променљивих што лоше утиче на резултате класификације. Због тога се природа променљивих овде испитивала помоћу корелационе матрице, а моделовала помоћу оператора минимум у складу са ИБА.

Табела 17. Резултати класификације за ЛК2 за нормализацију N_3 и поделу 50%-50%

	ЛК2.1		ЛК2.2		ЛК2.3	
	Тачност	Вар.	Тачност	Вар.	Тачност	Вар.
1-NN	0.7372	0.0028	0.7456	0.0028	0.8789	0.0019
3-NN	0.7958	0.0024	0.7933	0.0021	**0.8889**	0.0024
5-NN	0.8103	0.0017	0.8078	0.0023	0.8753	0.0021
7-NN	0.8228	0.0024	0.8144	0.0024	0.8481	0.0028
9-NN	**0.8244**	0.0018	**0.8219**	0.0018	0.8411	0.0027
11-NN	0.8211	0.0023	0.8161	0.0021	0.8178	0.0025

У Табели 17 приказани су резултати за нормализациону функцију N_3. Подела узорка ни овде није имала значајан утицај. Тестирање је вршено за k од 1 до 11 суседа, након чега је тачност кренула полако да опада за већину класификатора. Као што је и очекивано, линеарни класификатори ЛК2.1 није постигао високу тачност. Ни нелинеарни класификатор ЛК2.2 није постигао боље резултате, вероватно због високе корелисаности у подацима. Међутим, класификатор на бази експертске функције ЛК2.3 је постигао значајно повећање тачности класификације у односу на претходне класификаторе. Најбоља тачност приближно 89% је постигута за модел 3-NN који узима у обзир три најближа суседа. Из угла постигнуте тачности класификатор ЛК2.3 са k=3 је надмашио Алтманове Z скор моделе.

Табела 18. Просечна матрица конфузије за 3-NN за ЛК2.3

Стварно/ предвиђено	Активна	Банкрот
Активна	17,09 (47,5%)	0,91 (2,5%)
Банкрот	3,09 (8,6%)	14,91 (41,4%)

У матрици конфузије (Табела 18) налазе се просечне вредности за 100 понављања. На тест узорку од 36 компанија, погрешно су класификоване четири компаније у просеку. Грешка типа 2 даје информацију о броју компанија које су активне, а класификоване да су банкротирале што је у овом случају просечно једна компанија. Грешка типа 1 је значајнија са аспекта доносиоца одлуке јер говори о броју банкротиралих компанија које су класификоване као активне, што је у овом случају 3 компаније у просеку.

Иако је укупна тачност класификације мања, занимљиво је да је код Алтманових Z скор модела обрнут случај – мања је грешка типа 1, а већа грешка типа 2.

8.3.3. Резултати експеримента ЛК3

ЛК3 класификатор је заснован на најближем центроиду и конструише се помоћу параметарске ИБА мере сличности и аритметичке средине као оператора логичке агрегације. Овде је извршено тестирање ИБА мере сличности са једним и са два параметра.

ЛК3.1 користи једнопараметарску ИБА меру као основ за мерење сличности. Овај модел је упоређен са класификатором базираним на једнопараметарској ГЛСГП при подели узорка 50%-50% и коришћена је N_3 нормализациона функција. Резултати поређења су приказани у Табели 19.

Табела 19. Резултати класификације за ЛК3.1 на проблему предвиђања банкротства

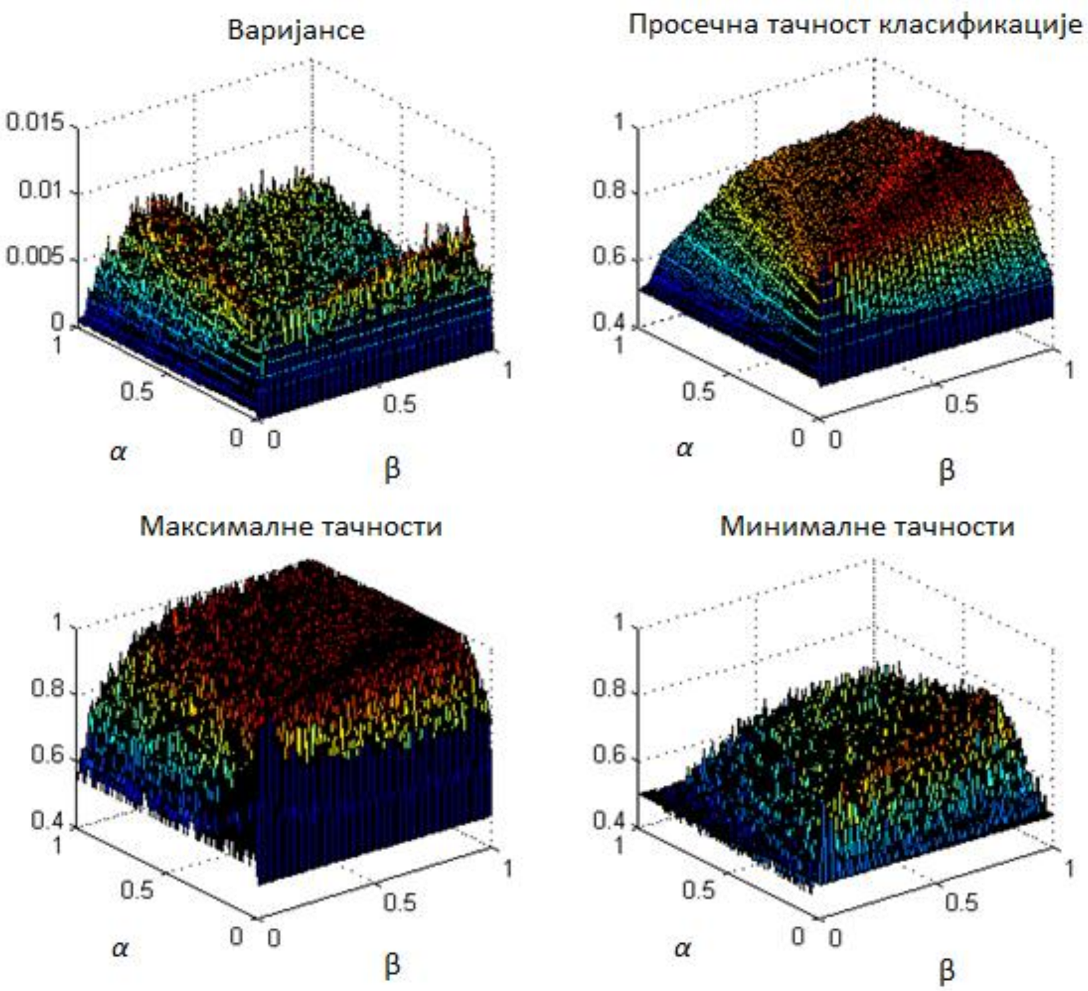
Нормализација	$S_{GLSGP}^{p,1}$			S_{IBA}^{α}		
	Тачност	Вар.	p	Тачност	Вар.	α
N_3	0,8811	0,0024	2,50	**0.8900**	0.0013	0,35

Приближни резултати су добијени за обе параметарске мере сличности.

ЛК3.2 користи двопараметарску ИБА меру као основ за мерење сличности. Овај модел је упоређен са класификатором базираним на двопараметарској ГЛСГП при подели узорка 50%-50% и коришћена је N_3 нормализациона функција. Резултати поређења су приказани у Табели 20.

Табела 20. Резултати класификације за ЛК3.2 на проблему предвиђања банкротства

Нормализација	$S_{GLSGP}^{p,m}$			$S_{IBA}^{\alpha,\beta}$		
	Тачност	Вар.	$(p;m)$	Тачност	Вар.	$(\alpha;\beta)$
N_3	0,8891	0,0028	(2,75; 1,25)	**0.8903**	0.0016	(0,29; 0,80)



Слика 18. Зависност тачности од параметара код ИБА мере сличности – предузећа

Графици зависности тачности и варијансе од вредности параметара код ГЛСГП мере у случају предузећа приложени су у Прилогу Д.

Табела 21. Просечна матрица конфузије за ЛК3.2

Стварно/ предвиђено	Активна	Банкрот
Активна	17,59 (48,86%)	0,41 (1,13%)
Банкрот	3,54 (9,83%)	14,46 (40,17%)

Као и код ЛК2 добијена је приближна тачност и структура матрице конфузије (Табела 21). Занимљиво је да се добио приближно исти резултат за класификатор са обичном ИБА мером сличности и експертском логичком функцијом агрегације и за класификатор са параметарском ИБА мером сличности и обичном аритметичком средином као функцијом агрегације.

Природно је било и тестирати комбинацију експертске функције агрегације заједно са параметарском ИБА мером сличности за коју се параметар оцењује из података. Међутим, није дошло до очекиваног повећања тачности класификације (Табела 22), па ће се у будућем истраживању подешавати и експертска функција агрегације сличности.

Табела 22. Резултати класификације за параметарску ИБА меру и експертску функцију – предузећа

Нормализација	ЛК2.3 са S_{IBA}^{α}			ЛК2.3 са $S_{IBA}^{\alpha,\beta}$		
	Тачност		α	Тачност	Вар.	$(\alpha;\beta)$
N_3	0,8828	0,0017	0,42	**0,8972**	0,0016	(0,37; 0,47)

8.4. Закључак експеримента

Основни циљ је био да се тестира способност логичких класификатора да унапреде перформансе класификације са аспекта тачности, а пре свега транспарентности конструисањем експертских логичких функција помоћу логичке агрегације као веома дескриптивног оператора агрегације. При развијању модела за предвиђање банкрота, интерпретација резултата и транспарентност алгоритма су веома важне за финансијске експерте *(Larose, 2014).* То је један од основних разлога зашто је тестирање извршено управо на проблему предвиђања банкрота за средња предузећа у Србији.

Најбољи резултати су постигнути са класификатором *k*-најближих суседа заснованог на непараметарској ИБА мери сличности и експертској логичкој агрегација (ЛК2), као и са класификатором најближи центроид заснованог на параметарској ИБА мера сличности и аритметичкој средини (ЛК3). Иако је први експертски, а други класификатор научен из података занимљиво је да су оба дала приближно исту тачност. Међутим, резултати класификације нису унапређени комбиновањем експертске функције агрегације и параметарске мере сличности која се оцењује из података. Као будуће истраживање, разматраће се могућности да се истовремено подешавају и параметарска мера сличности и логичка функција агрегације.

9. ЗАКЉУЧАК

Појам сличности има веома значајну улогу у људском расуђивању и учењу. Спроведена су многа истраживања која покушавају да моделују људско расуђивање засновано на сличности *(Chen, 2009).* Класична логика није се показала као одговарајућа за моделовање људског резоновања јер не може да опише „несавршено" знање у облику неодређености, неизвесности, непотпуности, итд. *(Minsky, 1985; McDermott 1987)*. Последњих деценија дошло је до експанзије логичких формализама као што су вишевредносне логике, укључујући фази логику које имају способност да моделују „несавршено" знање.

Докторска дисертација садржи целовит приказ области на којима почива предложени логички приступ моделовању сличности. Прегледом литературе утврђено је да су највише изучавани метрички приступи сличности. Њихова основна замерка огледа се у чињеници да метричка својства не морају бити испуњена у практичној примени, као и да ограничавају доменске експерте у моделовању сличности. Поред тога, постоје истраживања која су показала да мере поређења која не испуњавају метричке аксиоме дају подједнако добре резултате у практичној примени *(e.g. Saastamoinen, 2006).* Због тога је, у овом истраживању посебна пажња посвећена критичкој анализи метричких приступа. На основу прегледа стања у досадашњим истраживањима, уочена је потреба за увођењем модела сличности који пружају интерпретацију доносиоцима одлуке и који могу да унапреде постојеће системе за подршку одлучивању са аспекта транспарентности.

Први део истраживања се односио на увођење ИБА еквиваленција као интерпретабилне мере сличности. Дато је њено основно тумачење, графички приказ и показано је да она задовољава сва математичка својства мере сличности.

Такође, апликативна вредност ИБА мере сличности је илустрована на примерима закључивања на основу случајева и моделовања консензуса у групном одлучивању.

На основу добрих резултата које су дале постојеће параметризоване логичке мере поређења природан следећи корак је био увођење параметра код ИБА логичке мере сличности која представља централну тему овог истраживања. Развијене су нове параметарске ИБА мере сличности које се за карактеристичне вредности параметара своде на већ постојеће логичке мере сличности: минимум, производ и Лукашијевич фази би-импликацију. Једна од основних препрека за примену параметарских модела је недостатак интерпретације. У случају параметарских ИБА мера сличности дато је јасно тумачење параметара, као и графичка приказ. За потребе евалуације дефинисаних параметарских мера сличности уведена је логичка класификацији заснована на сличности. На примеру медицинских података експериментално је потврђено да увођење параметара код ИБА мере сличности унапређује резултате класификације са аспекта тачности, времена извршавања и интерпретације у односу на постојеће параметарске мере поређења.

У наставку истраживања циљ је био да се покаже транспарентност логичке класификације засноване на ИБА. ЛК је примењена на проблем предвиђања банкрота предузећа код кога је управо транспарентост методе јако важна за доносиоца одлука. Експериментално је потврђено да логички класификатори са интерпретабилном мером и/или експертском логичком функцијом агрегације доприносе тачности класификације и имају семантичку вредност за доносиоца одлуке.

У наставку је дат осврт на постављене хипотезе, доприносе и могућности даљег истраживања.

9.1. Осврт на постављене хипотезе и остварене доприносе

Ово поглавље је део закључка који представља осврт на постављене хипотезе и остварене доприносе у оквиру докторског рада.

Основне хипотезе:

- Логички приступ моделује сличност на начин који је близак људском расуђивању.

  Вишевредносне релације еквиваленције и импликације давно су уведене као основа за природан начин мерења сличности. Логички приступ заснован на ИБА еквиваленцији нуди нови начин посматрања сличности, где се поред поседовања особине објекти упоређују и према одсуству неке особине. Поред тога, као последицу дескриптивности оператора логичке агрегације могуће је конструисати и експертске логичке функције блиске доносиоцима одлука. Транспарентност и интерпретабилност логичког приступа за доносиоца одлука показане су на илустративним примерима закључивања на основу случајева, моделовања консензуса и на примеру класификације над реалним подацима.

- Алати за подршку одлучивању могу се унапредити увођењем логичких мера сличности.

  У експериментима је емпиријски потврђено да теоријски уведене непараметарске и параметерске ИБА мере сличности заједно са логичком агрегацијом унапређује класификаторе из угла тачности, времена извршавања, интерпретације, дескриптивности и транспарентности за доносиоца одлуке. Јасна интерпретација логичког приступа моделовању сличности помоћу ИБА показана је на илустративним примерима консензуса и закључивања на основу случајева.

Помоћне хипотезе:

- Логичке релације импликације и еквиваленције могу се користити за мерење сличности између објеката.

  Прегледом литературе утврђено је да постоји значајан број вишевредносних релације еквиваленције и импликације које су се показале као успешне при моделовању сличности.

- Релација еквиваленције заснована на интерполативној Буловој алгебри може се дефинисати као мера сличности.

  Провером својстава ненегативности, симетричности и ограничености са горње стране показано је да ИБА еквиваленција испуњава услове за меру сличности.

- Мера сличности заснована на ИБА еквиваленцији може се уопштити увођењем параметра - параметарска ИБА еквиваленција.

  Показано је да се за карактеристичне вредности параметара параметарска ИБА еквиваленција своди се на друге логичке мере сличности: минимум, производ и Лукашијевич фази би-импликацију. Увођењем параметара проширује се површ коју покрива ИБА мера сличности што је илустровано и графички за различите случајеве.

- Сличност се може моделовати помоћу псеудо-логичке агрегације и мера сличности заснованих на ИБА.

  Помоћу ИБА мере сличности и логичке агрегације конструисани су различити класификатори засновани на сличности који су дали добре резултате за предвиђање банкротства предузећа. Такође, на овај начин моделована је сличност и код закључивања на основу случајева.

- Логичке мере сличности засноване на ИБА могу се користити за моделовање консензуса.

  На илустративном примеру показано је да моделовање консензуса коришћењем ИБА омогућава адекватну интерпретацију консензуса и одговарајући степен слагања експерата.

- На основу ИБА логичких мера сличности могу се образовати различити класификатори.

  У експериментима су конструисани различити класификатори засновани на параметарским/непараметарским мерама сличности и логичкој агрегацији који су дали добре резулате класификације и имају семантичку вредност за доносиоца одлука.

Резултати истраживања у докторској дисертацији садрже следеће научне доприносе:

- Извршена је критичка анализа математичких својстава постојећих метричких приступа за мерење сличности.
- Дат је детаљан преглед постојећих логички заснованих мера сличности, као и параметризованих мера поређења.
- Дефинисана је нова интерпретабилна логичка мера сличности заснована на ИБА еквиваленцији.
- ИБА мера сличности је потврђена на илустративним примерима закључивања на основу случајева и консензуса, као и на проблему класификације, где је извршено и поређење са другим постојећим логички заснованим мерама сличности.
- Извршена је параметризација ИБА мере сличности тако да се друге логичке мере сличности засноване на минимуму, производу и Лукашијевич фази би-импликацији могу добити као подслучај параметарске ИБА еквиваленције. Објашњена је мотивација за увођење параметара и дато основно тумачење.
- Дефинисана је нова параметарска ИБА мера сличности.
- Ради валидације параметарских и непараметарских ИБА мера сличности уведен је нови логички приступ класификацији који чува семантичку информацију за доносиоца одлуке помоћу интерпретабилних ИБА мера сличности, логичке агрегације као веома дескриптивног оператора агрегације и два начина поређења објеката: према особини и изведеној особини.
- Вредност параметарске ИБА мере сличности је потврђена конструисањем класификатора заснованим на сличности где су параметри научени из података. Валидација је извршена са аспекта унапређења тачности и брзине извршавања на примеру познатих медицинских података који се користе за класификацију.
- Логички приступ класификацији је примењен на проблему предвиђања банкротства предузећа у Србији. Остварени су добри резултати класификације коришћењм параметарске ИБА мере сличности, али и

креирањем експертских функција агрегације сличности чиме је потврђена транспарентност приступа.

- Друштвени допринос се огледа у могућности примене предложених приступа у различитим областима.

9.2. Могући правци будућег истраживања

У току истраживања посебно су се издвојила два правца будућег истраживања:

- Проширење теоријског оквира параметарских ИБА мера сличности и домена примене и
- Унапређење класификатора заснованих на логичком ИБА приступу моделовања сличности.

Први правац се односи на потенцијално проширење дефиниције параметарских ИБА мера сличности. Неке од могућности већ су разматране у поглављу 5.3. а односе се на проширење опсега параметра на интервал [-1,1], а тиме потенцијално и на промену интерпретације.

Поред тога, параметарске мере сличности добро су се показале код проблема класификације патерна где је постојала могућност да се одбаце неки патерни као недовољно поуздани *(Le Capitaine & Frelicot, 2009a)*. У ту сврху размотриће се увођење мере неодређености засноване на ИБА параметарским мерама сличности.

У оквиру другог правца развоја, истраживање би се бавило унапређењем перформанси логички заснованих класификатора. Природан следећи корак је испитивање да ли би редукција димензионалности података унапредила перформансе класификације. Поред тога, једна од најбитнијих ствари за унапређење перформанси алгоритма је увођење одговарајућих оптимизатора при учењу параметара из података по угледу на (*Luukka & Leppalampi, 2006; Saastamoinen, 2006; Koloseni, 2015*).

Поред оцене параметара код мера сличности, разматраће се могућност да се почетно задате експертске логичке функције агрегације подесе из података, или да се комплетна структура оператора агрегације за неки проблем одреди из података.

Да би се реално сагледале перформансе логички заснованих класификатора неопходно је извршити поређење са успешним класификаторима не само из групе алгоритама са лењим учењем, већ и са другим моделима са класичним учењем, са линеарним класификаторима, нелинеарним неуронским мрежама, затим транспарентним методама као што су стабла одлучивања, итд.

Са аспекта реалне примене класификатора важно је модификовати функције нормализације (нпр. софтмакс методом омекшавања горње границе логистичком функцијом) како би имале способност да трансформишу и оне улазне вредности које излазе ван предефинисаног опсега код стандардних нормализационих функција. Занимљив је и приступ тражења оптималне мере сличности за сваки атрибут појединачно из групе предефинисаних мера сличности *(Koloseni, 2015).*

10. ПРИЛОЗИ

Прилог А – Резултати класификације за непараметарске логичке мере

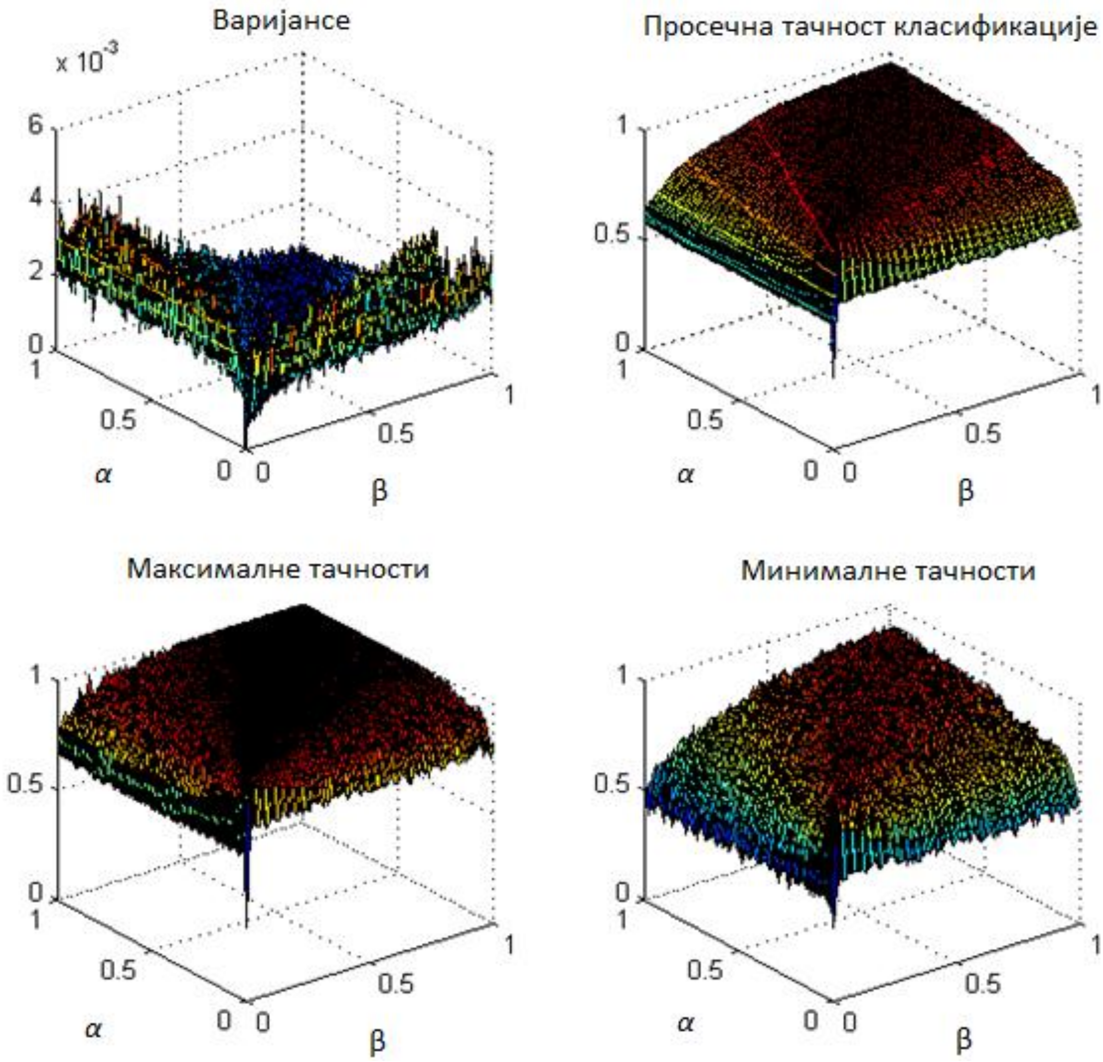
Табела 23. Резултати класификације за непараметарске логичке мере за нормализацију N_1 и поделу 70%-30%

	S_M		S_P		S_L		S_{IBA}	
	Тачност	Вар.	Тачност	Вар.	Тачност	Вар.	Тачност	Вар.
Iris	0,3333	0,0001	0,6727	0,0036	0,9256	0,0013	0,9289	0,0018
Wine	0,4196	0,0011	0,3918	0,0030	0,9638	0,0004	0,9607	0,0005
Dermatology	0,1867	0,0003	0,1863	0,0005	0,9579	0,0003	0,9595	0,0002
PIMA	0,3527	0,0000	0,4330	0,0004	0,7262	0,0007	0,7263	0,0007
BUPA	0,4264	0,0001	0,4503	0,0014	0,5458	0,0025	0,5360	0,0021
BCW1	0,9120	0,0003	0,9140	0,0003	0,9442	0,0002	0,9424	0,0002
WDBC	0,5680	0,0006	0,7741	0,0010	0,9419	0,0003	0,9408	0,0003
Thyroid	0,8446	0,0002	0,7737	0,0025	0,9417	0,0006	0,9429	0,0007

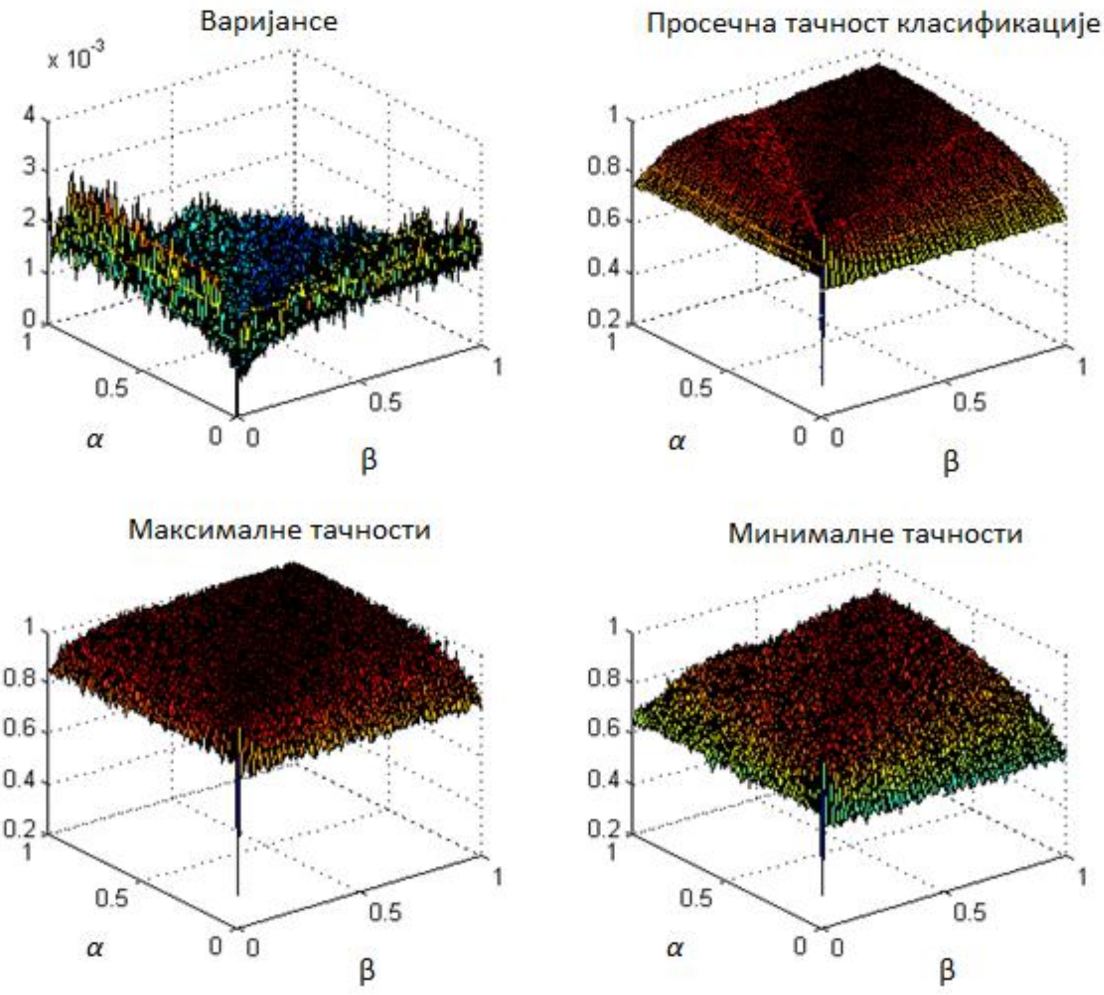
Табела 24. Резултати класификације за непараметарске логичке мере за нормализацију N_2 и поделу 70%-30%

	S_M		S_P		S_L		S_{IBA}	
	Тачност	Вар.	Тачност	Вар.	Тачност	Вар.	Тачност	Вар.
Iris	0,5011	0,0052	0,7493	0,0030	0,9560	0,0008	0,9500	0,0007
Wine	0,3273	0,0000	0,4125	0,0021	0,9425	0,0007	0,9475	0,0006
Dermatology	0,1875	0,0003	0,1886	0,0005	0,9584	0,0003	0,9562	0,0003
PIMA	0,3515	0,0000	0,4335	0,0004	0,7359	0,0007	0,7311	0,0006
BUPA	0,4231	0,0000	0,4327	0,0005	0,5476	0,0022	0,5555	0,0020
BCW1	0,9115	0,0003	0,9276	0,0002	0,9429	0,0001	0,9418	0,0002
WDBC	0,5681	0,0006	0,8137	0,0006	0,9375	0,0002	0,9379	0,0003
Thyroid	0,7665	0,0004	0,7777	0,011	0,9437	0,0008	0,9380	0,0008

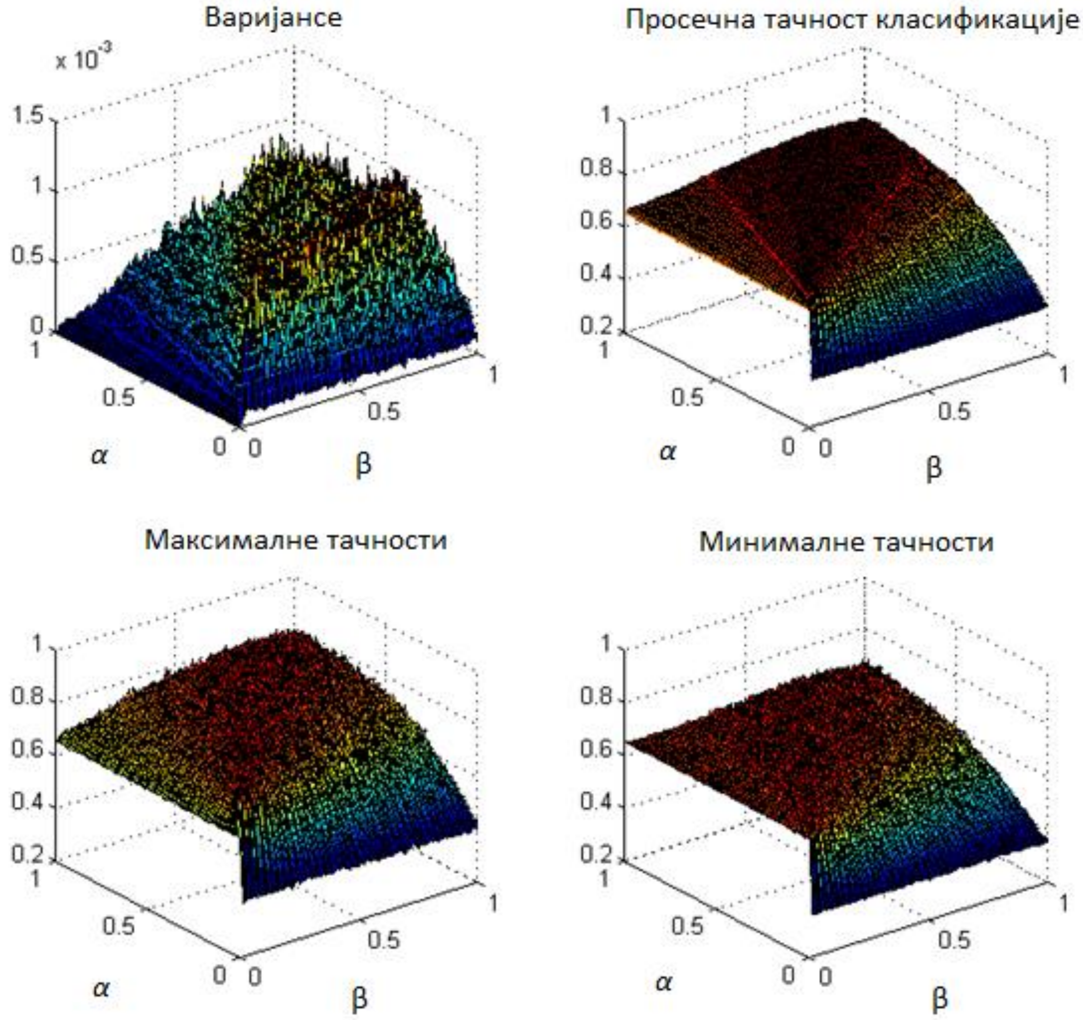
Прилог Б - Зависност тачности класификације на медицинским скуповима податка од параметара код ИБА мере сличности



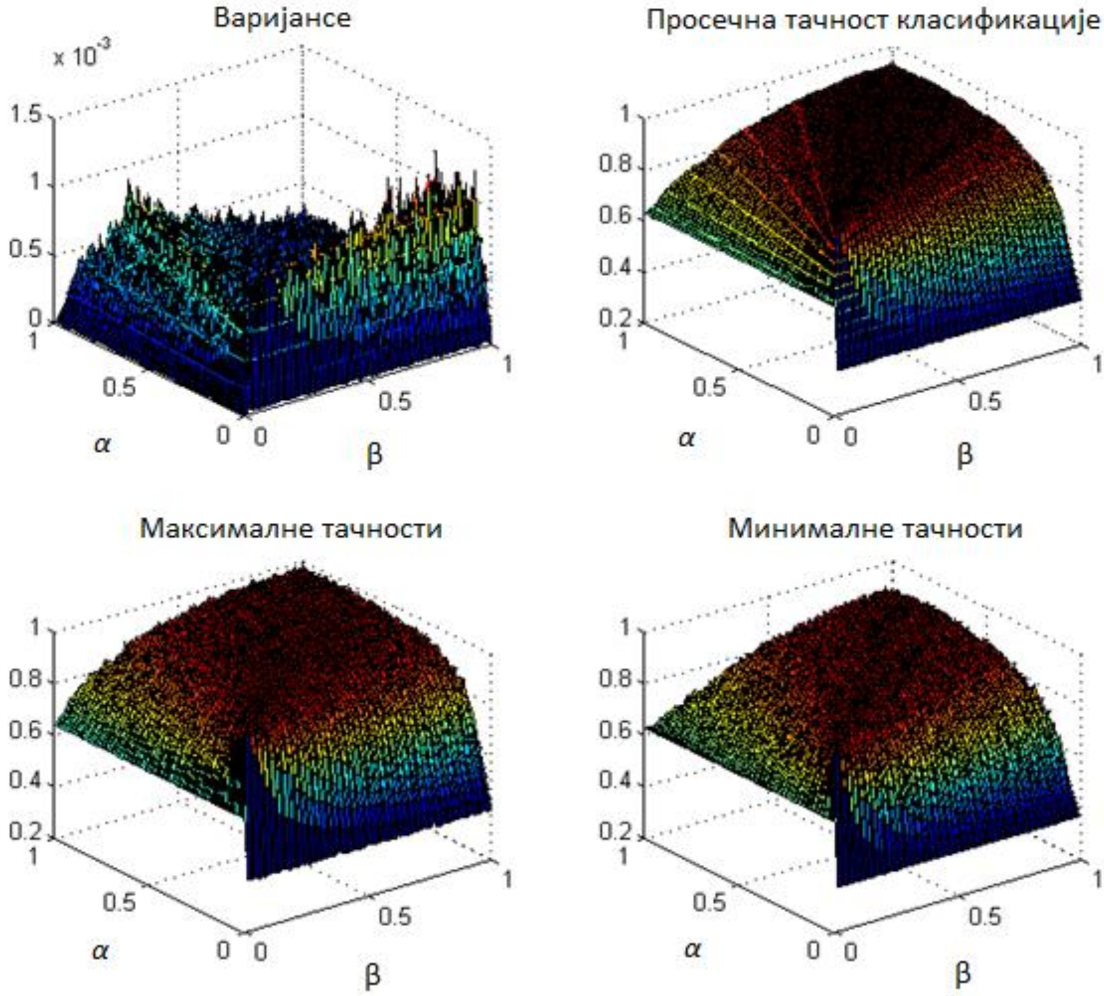
Слика 19. Зависност тачности од параметара код ИБА мере сличности – Iris



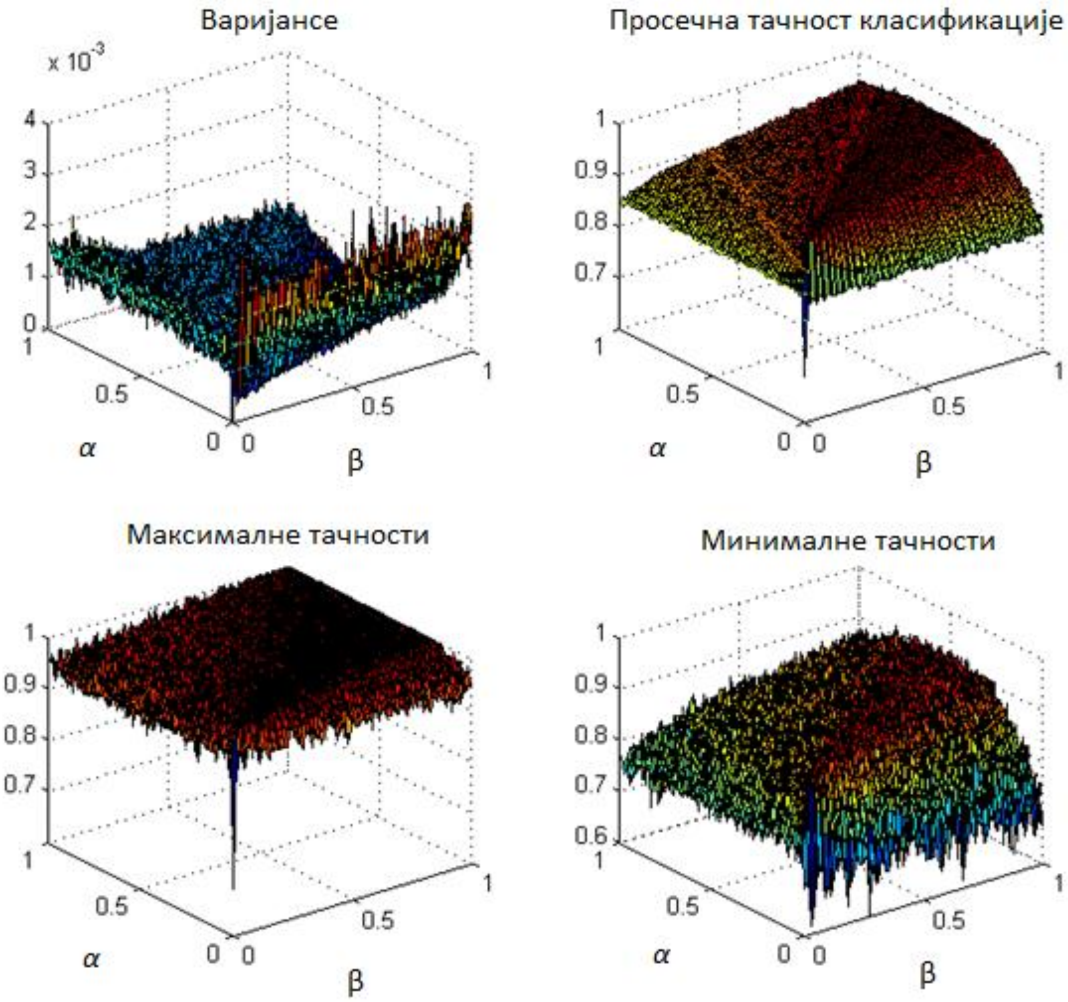
Слика 20. Зависност тачности од параметара код ИБА мере сличности – Wine



Слика 21. Зависност тачности од параметара код ИБА мере сличности – PIMA

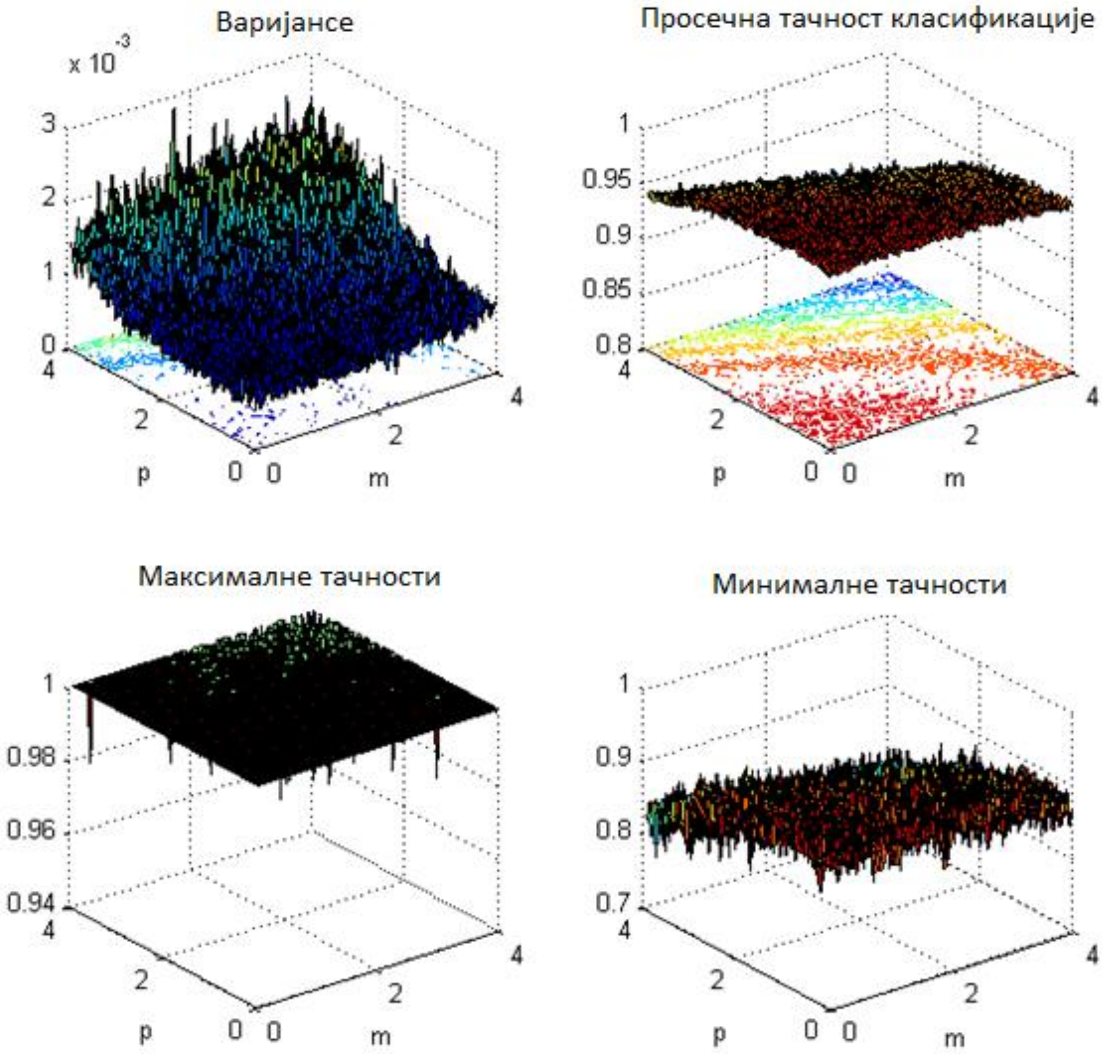


Слика 22. Зависност тачности од параметара код ИБА мере сличности – WDBC

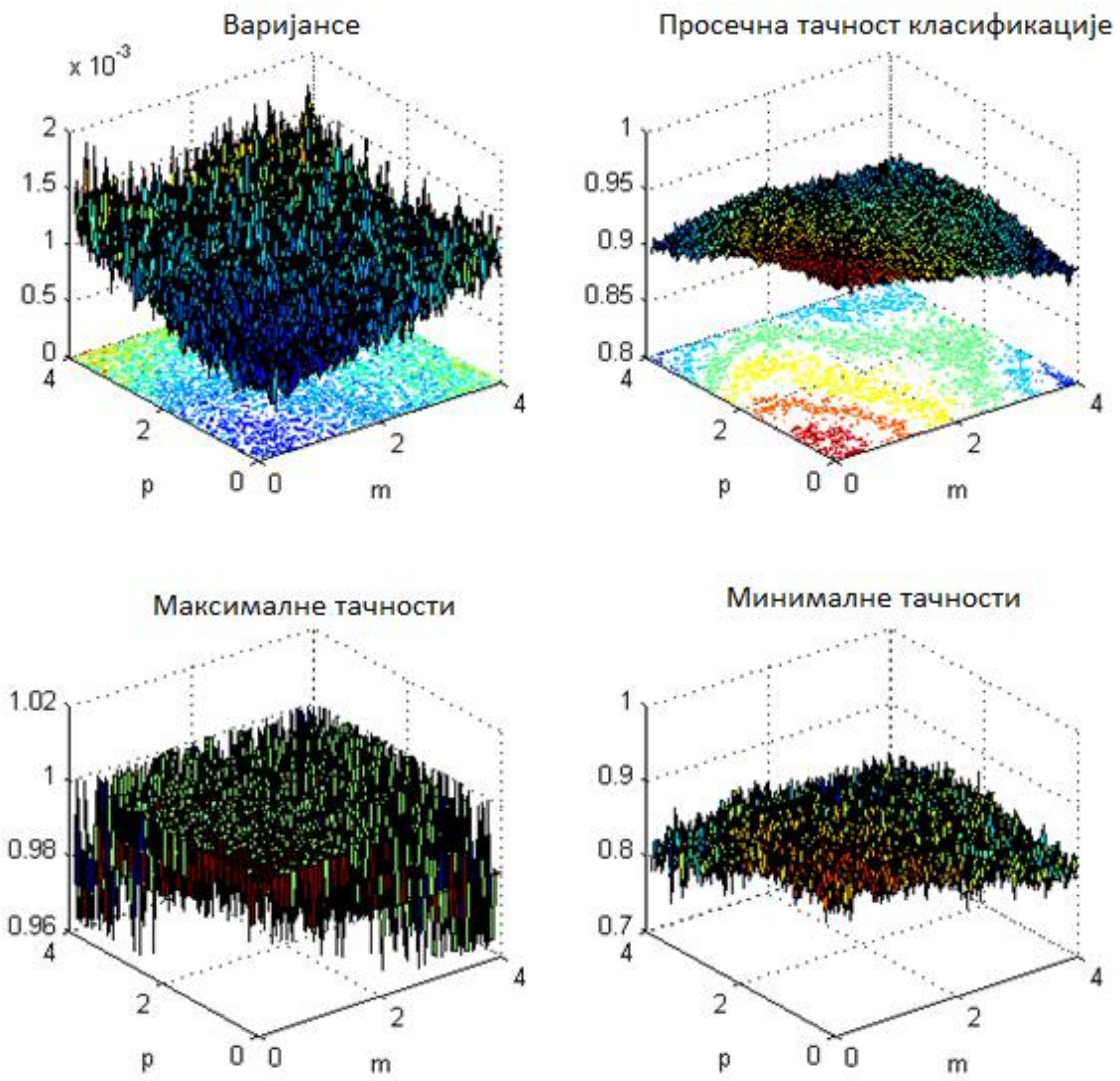


Слика 23. Зависност тачности од параметара код ИБА мере сличности – Thyroid

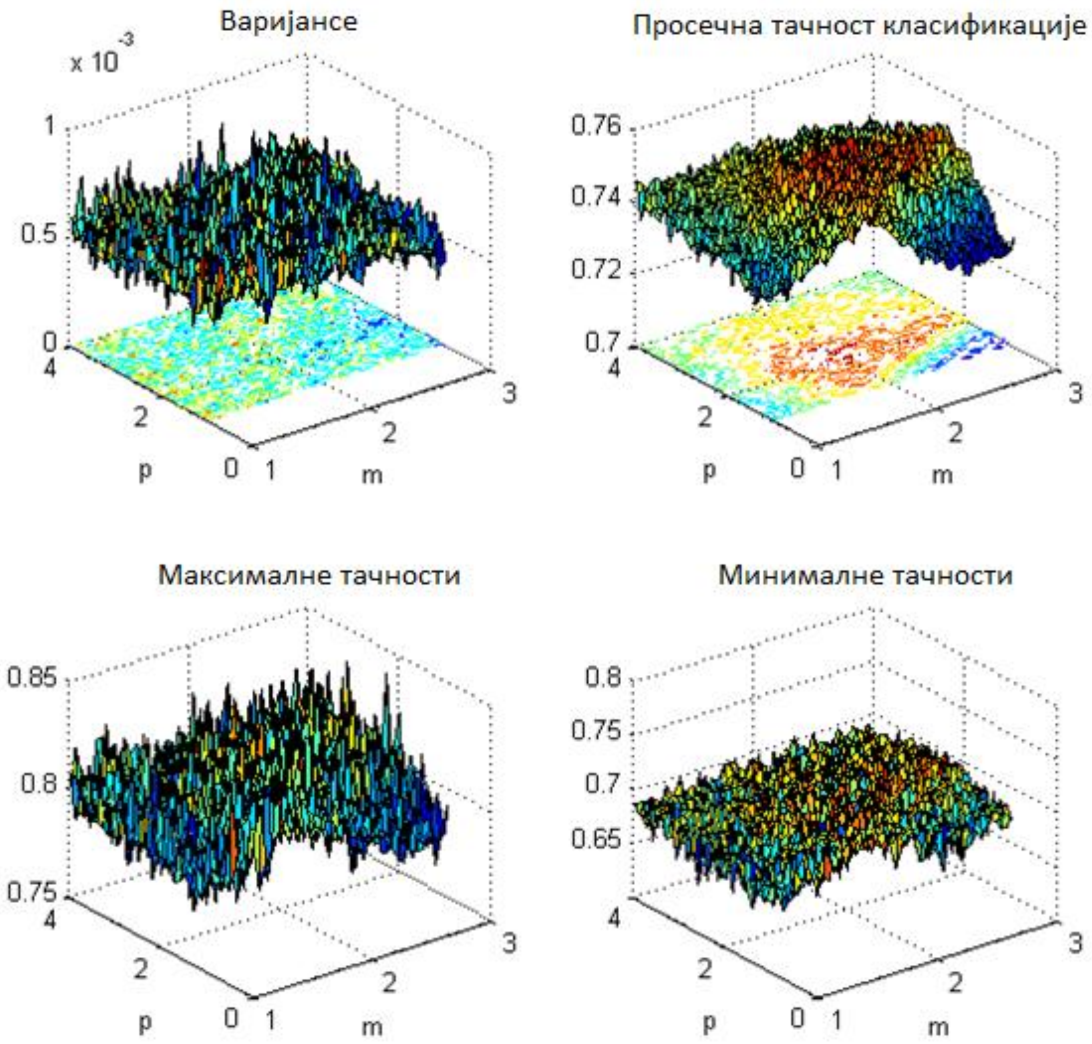
Прилог В - Зависност тачности класификације на медицинским скуповима податка од параметара код ГЛСГП мере сличности



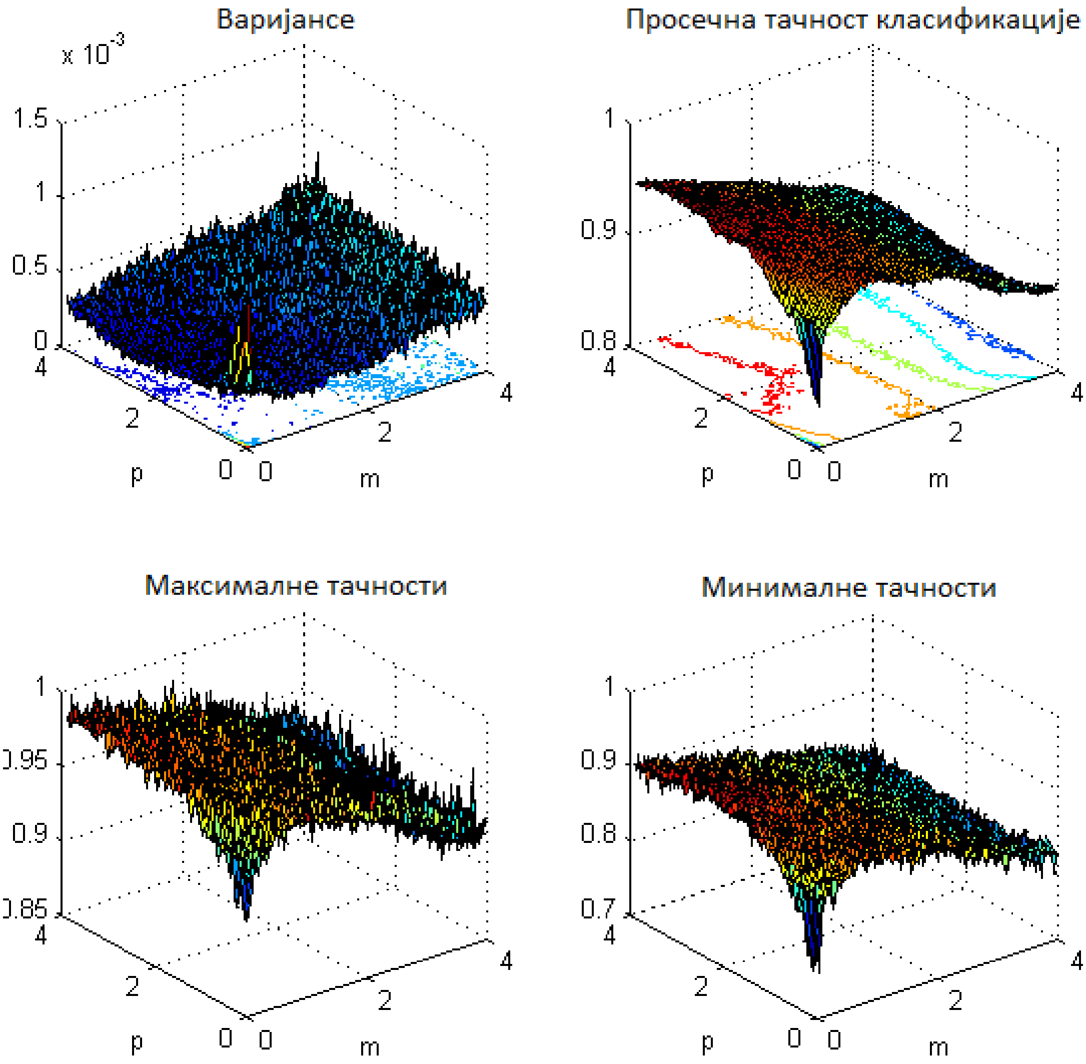
Слика 24. Зависност тачности од параметара код ГЛСГП – Iris



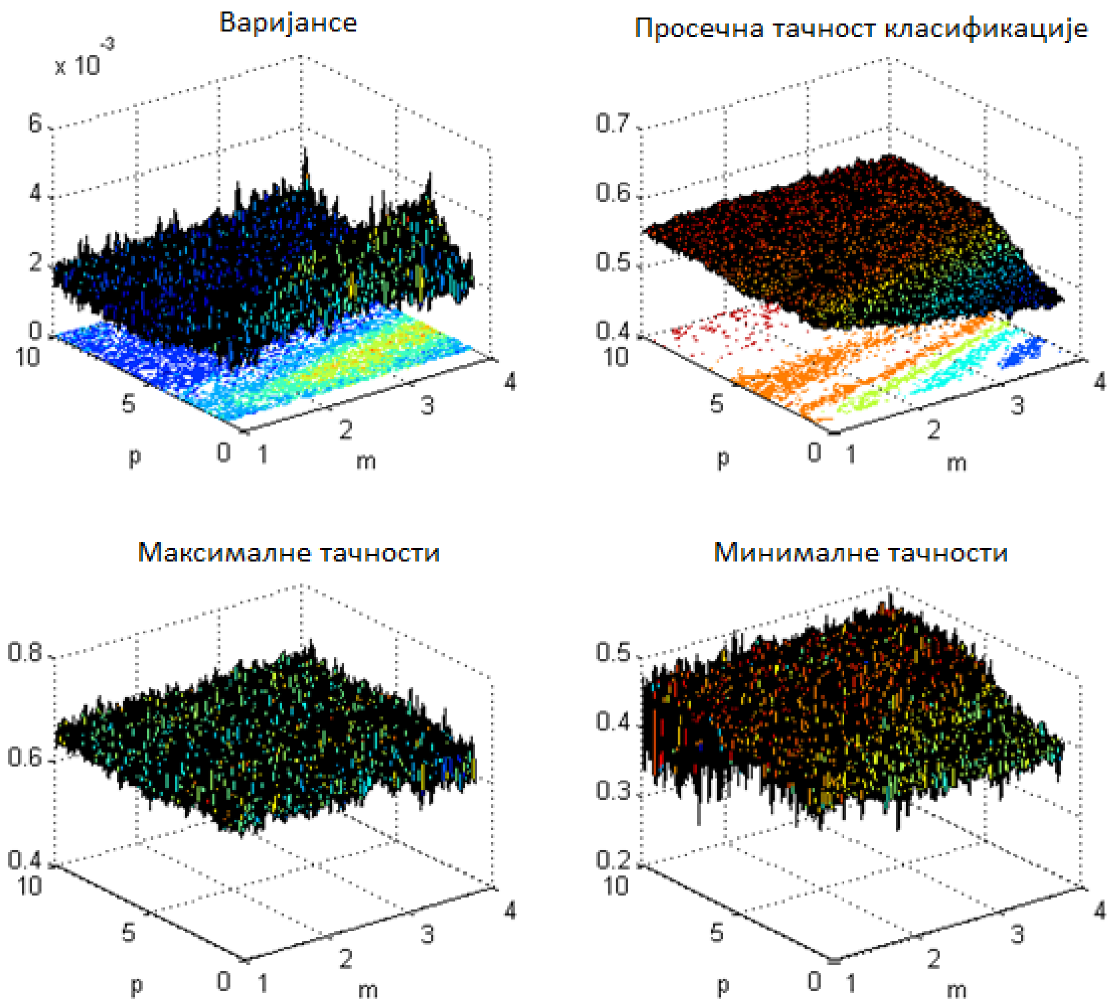
Слика 25. Зависност тачности од параметара код ГЛСГП – Wine



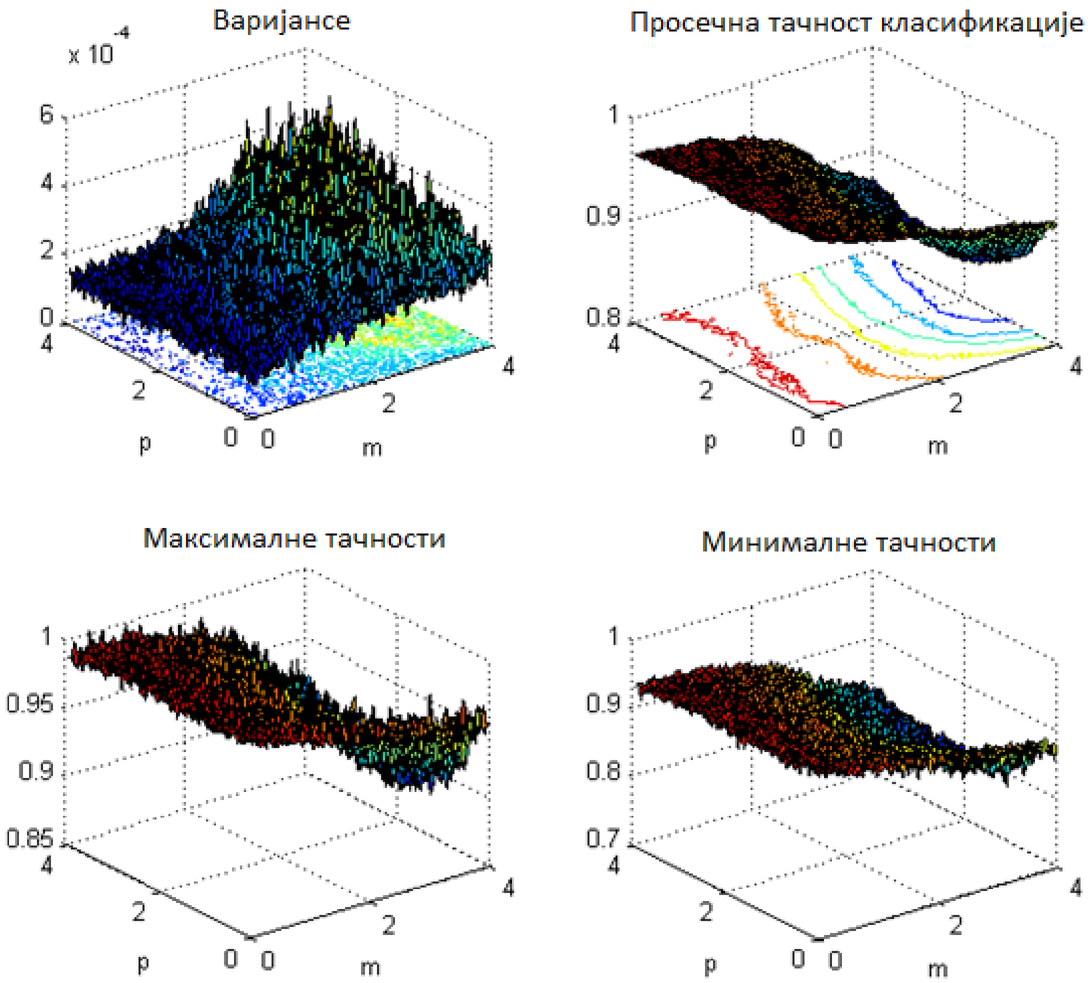
Слика 26. Зависност тачности од параметара код ГЛСГП – PIMA



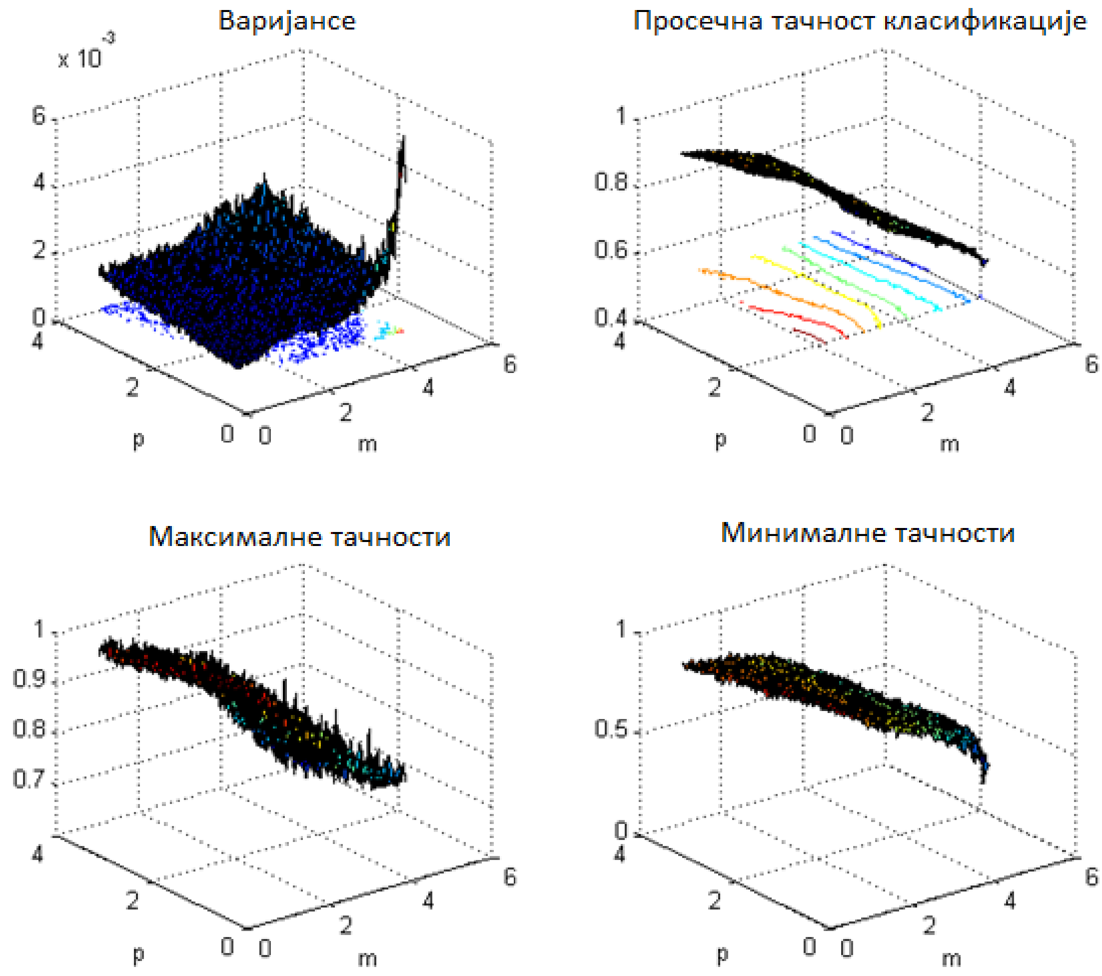
Слика 27. Зависност тачности од параметара код ГЛСГП – WDBC



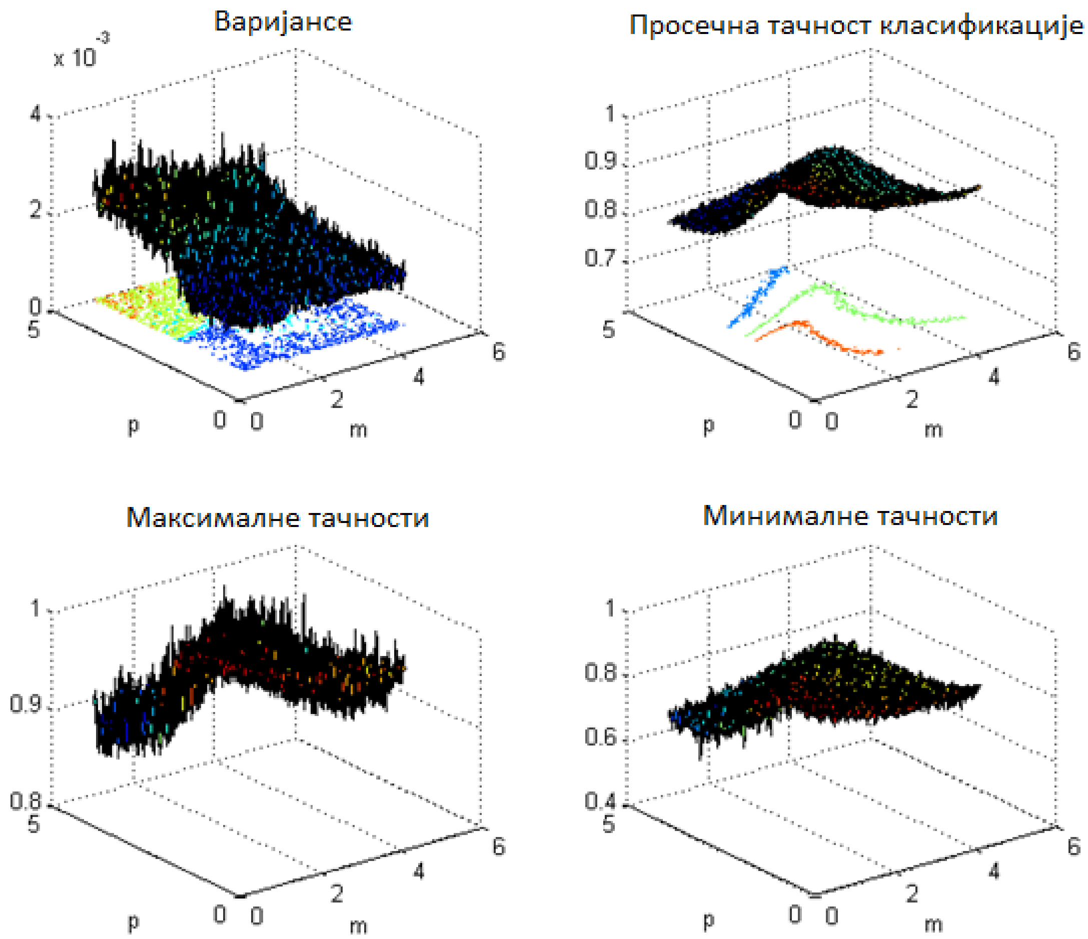
Слика 28. Зависност тачности од параметара код ГЛСГП – BUPA



Слика 29. Зависност тачности од параметара код ГЛСГП – BCW1



Слика 30. Зависност тачности од параметара код ГЛСГП – Dermatology



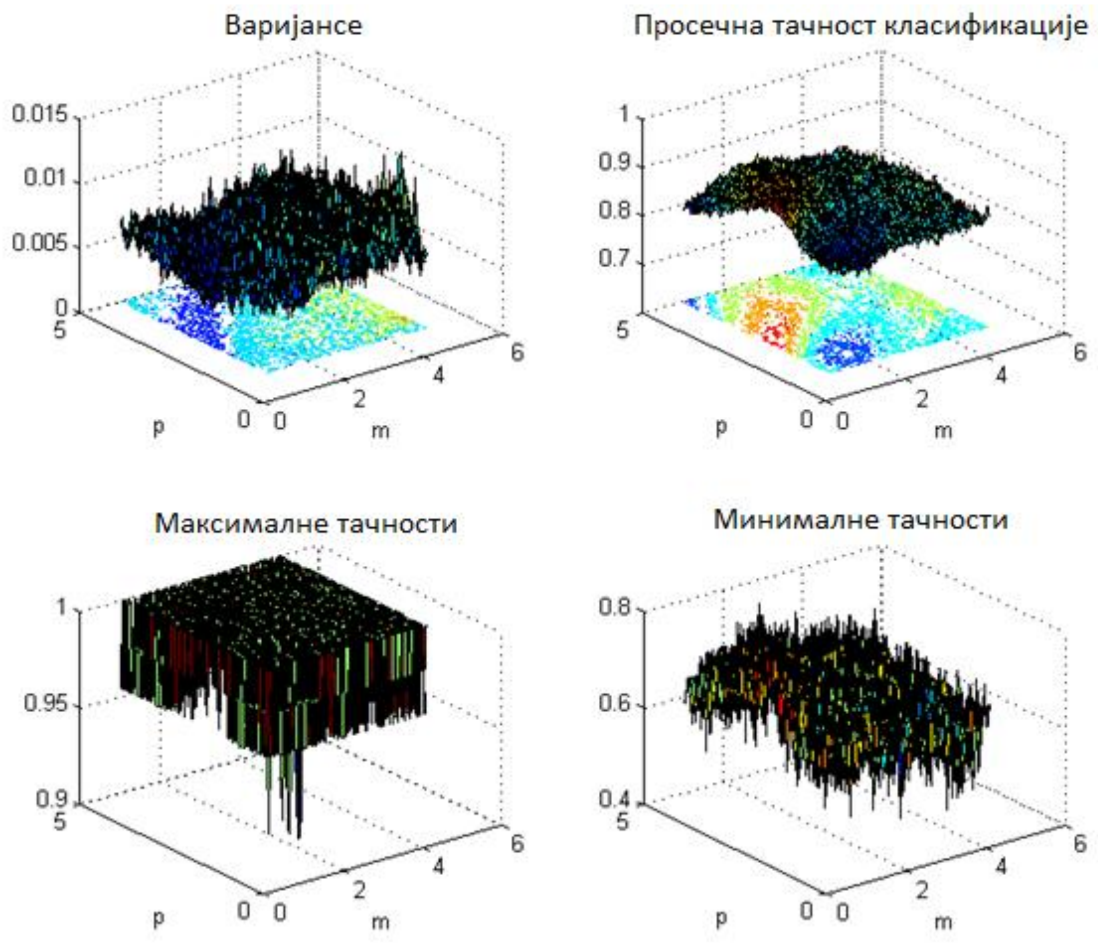
Слика 31. Зависност тачности од параметара код ГЛСГП – Thyroid

Прилог Г – Резултати класификације за параметарску ИБА меру сличности са вредностима параметра са интервала од -1 до 1

Табела 25. Резултати класификације за параметарску ИБА меру за N_3 и поделу 70%-30% – за α од -1 до 1

	S_{IBA}		$S_{IBA}^{\alpha[-1,1]}$			S_{IBA}^{α}	$S_{IBA}^{\alpha,\beta}$
	Тачност	Вар.	Тачност	Вар.	α	Тачност	Тачност
Iris	0,9570	0,0007	0,9600	0,0007	0,95	0,961	0,9633
Wine	0,9420	0,0007	0,9458	0,0006	0,95	0,9438	0.9480
Dermatology	0,9572	0,0003	0,9766	0,0002	0,55	0,9775	0,9773
PIMA	0,7350	0,0009	0,7350	0,0009	1,00	0,7283	0,7442
BUPA	0,5494	0,0021	0,5828	0,0022	-0,15	0,5776	0,5833
BCW1	0,9414	0,0001	0,9737	0,0001	0,35	0,9737	0,9751
WDBC	0,9324	0,0002	0,9370	0,0003	0,8	0,9387	0,9401
Thyroid	0,9434	0,0008	0,9651	0,0004	0,45	0,9615	0,9663

Прилог Д – Тачности класификације предузећа са аспекта банкротства у зависности од параметара коришћењем ГЛСГП мере сличности



Слика 32. Зависност тачности од параметара код ГЛСГС – предузећа

11. ЛИТЕРАТУРА

Aeberhard, S., Coomans, D., & De Vel, O. (1992). The classification performance of RDA. *Dept. of Computer Science and Dept. of Mathematics and Statistics, James Cook University of North Queensland, Tech. Rep,* 92-01.

Aggarwal, C. C., & Zhai, C. (2012). A survey of text classification algorithms. In *Mining text data* (pp. 163-222). Springer US.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, *23*(4), 589-609.

Altman, E. I. (2000). Predicting financial distress of companies: revisiting the Z-score and ZETA models. *Stern School of Business, New York University*, 9-12.

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2016). Financial distress prediction in an international context: A Review and Empirical Analysis of Altman's z-score model. *Journal of International Financial Management & Accounting, 0:0,* DOI: 10.1111/jifm.12053.

Baczyński, M., & Jayaram, B. (2008a). (S, N)-and R-implications: A state-of-the-art survey. *Fuzzy Sets and Systems*, *159*(14), 1836-1859.

Baczyński, M., & Jayaram, B. (2008b). *Fuzzy Implications (Studies in Fuzziness and Soft Computing, Vol. 231)*. Springer Berlin Heidelberg.

Bandler, W., & Kohout, L. (1980). Fuzzy power sets and fuzzy implication operators. *Fuzzy Sets and Systems*, *4*(1), 13-30.

Bandyopadhyay, S., & Saha, S. (2012). *Unsupervised classification: similarity measures, classical and metaheuristic approaches, and applications*. Springer Science & Business Media.

Beliakov, G., Calvo, T., & James, S. (2014). Consensus measures constructed from aggregation functions and fuzzy implications. *Knowledge-Based Systems*, *55*, 1-8.

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, *18*(9), 509-517.

Bloch, I., & Maitre, H. (1995a, February). Fuzzy distances and image processing. In *Proceedings of the 1995 ACM symposium on Applied computing* (pp. 570-574). ACM.

Bloch, I., & Maitre, H. (1995b). Fuzzy mathematical morphologies: a comparative study. *Pattern recognition*, *28*(9), 1341-1387.

Bloch, I. (1999). On fuzzy distances and their use in image processing under imprecision. *Pattern Recognition*, *32*(11), 1873-1895.

Bouchon-Meunier, B., Rifqi, M., & Bothorel, S. (1996). Towards general measures of comparison of objects. *Fuzzy sets and systems*, *84*(2), 143-153.

Bourbaki, N. (2004). Theory of sets. In *Theory of sets* (pp. 65-129). Springer Berlin Heidelberg.

Callejas, C., Marcos, J., & Bedregal, B. R. C. (2012). On some subclasses of the Fodor-Roubens fuzzy bi-implication. In *Logic, Language, Information and Computation. Proceedings of WoLLIC 2012.* (pp. 206-215). Springer Berlin Heidelberg.

Chen, Y., Garcia, E. K., Gupta, M. R., Rahimi, A., & Cazzanti, L. (2009). Similarity-based classification: Concepts and algorithms. *Journal of Machine Learning Research*, *10*(Mar), 747-776.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, *13*(1), 21-27.

Cross, V. (2006, June). Tversky's parameterized similarity ratio model: a basis for semantic relatedness. In *NAFIPS 2006-2006 Annual Meeting of the North American Fuzzy Information Processing Society* (pp. 541-546). IEEE.

Cross, V. V., & Sudkamp, T. A. (2002). *Similarity and compatibility in fuzzy set theory: assessment and applications (Vol. 93)*. Springer Science & Business Media.

Cross, V., & Cabello, C. (1995, September). A mathematical relationship between set-theoretic and metric compatibility measures. In *Proceedings of Third International Symposium on Uncertainty Modeling and Analysis and Annual Conference of the North American Fuzzy Information Processing Society ISUMA – NAFIPS '95* (pp. 169-174). IEEE.

Ćirić, M., Ignjatović, J., & Bogdanović, S. (2007). Fuzzy equivalence relations and their equivalence classes. *Fuzzy Sets and Systems*, *158*(12), 1295-1313.

Coomans, D., Broeckaert, I., Jonckheer, M., & Massart, D. L. (1983). Comparison of multivariate discrimination techniques for clinical data-application to the thyroid functional state. *Methods of Information in Medicine*, *22*(2), 93-101.

Dasarathy, B. V. (1991). *Nearest neighbor ({NN}) norms:{NN} pattern classification techniques*. IEEE Computer Society Press, Los Alamitos.

De Cock, M., & Kerre, E. (2003). On (un) suitable fuzzy relations to model approximate equality. *Fuzzy Sets and Systems*, *133*(2), 137-153.

De Baets, B., & Mesiar, R. (1997). Pseudo-metrics and T-equivalences. *Journal of Fuzzy Mathematics*, *5*, 471-481.

De Baets, B., & Mesiar, R. (1998). T-partitions. *Fuzzy Sets and Systems*, *97*(2), 211-223.

De Baets, B., & Mesiar, R. (2002). Metrics and T-equalities. *Journal of Mathematical Analysis and Applications*, *267*(2), 531-547.

Demirci, M. (2003a). Indistinguishability operators in measurement theory, Part II: Construction of indistinguishability operators based on probability distributions. *International Journal of General Systems, 32*(5), 431-458.

Demirci, M. (2003b). On many-valued partitions and many-valued equivalence relations. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *11*(02), 235-253.

Demirci, M. (2003c). Representations of the extensions of many-valued equivalence relations. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *11*(03), 319-341.

Deza, M. M., & Deza, E. (2014). *Encyclopedia of distances*. Springer Berlin Heidelberg.

Di Nola, A., Pedrycz, W., & Sessa, S. (1988). Fuzzy relation equations with equality and difference composition operators. *Fuzzy Sets and Systems, 25*(2), 205-215.

Donald, H., Joseph, C., Don, C., David, G., & Steven, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology*, *101*(1), 116.

Dobrić, V., Kovačević, D., Petrović, B., Radojević, D., & Milošević, P. (2015). Formalization of human categorization process using Interpolative Boolean algebra. *Mathematical Problems in Engineering*, 2015, Article ID 620797.

Dragović, I., Turajlić, N., Radojević, D., & Petrović, B. (2014). Combining Boolean consistent fuzzy logic and AHP illustrated on the web service selection problem. *International Journal of Computational Intelligence Systems*, 7(1), 84-93.

Dubois, D., & Jaulent, M. C. (1987). A general approach to parameter evaluation in fuzzy digital pictures. *Pattern Recognition Letters, 6*(4), 251-259.

Eckhardt, A., Skopal, T., & Vojtáš, P. (2009). On fuzzy vs. metric similarity search in complex databases. In *Flexible Query Answering Systems* (pp. 64-75). Springer Berlin Heidelberg.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, *7*(2), 179-188.

Fix, E., & Hodges Jr, J. L. (1951). *Discriminatory analysis - Nonparametric discrimination: Consistency properties (Report No. ADA800276)*. California University Berkeley. Link: http://www.dtic.mil/get-tr-doc/pdf?AD=ADA800276. [Pristupljeno 17.05.2016].

Fodor, J. C., & Roubens, M. R. (1994). *Fuzzy preference modelling and multicriteria decision support* (Vol. 14). Springer Science & Business Media.

Formato, F., Gerla, G., & Scarpati, L. (1999). Fuzzy subgroups and similarities. *Soft Computing*, *3*(1), 1-6.

Fréchet, M. M. (1906). Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo* (1884-1940), 22(1), 1-72.

Friedman, J. H., Bentley, J. L., & Finkel, R. A. (1977). An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS)*, *3*(3), 209-226.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning: Data mining, inference and prediction* (Vol. 1). Springer Science & Business Media.

Gan, G., Ma, C., Wu, J., (2007). *Data clustering: theory, algorithms, and applications (Vol. 20)*. SIAM.

Godo, L., & Rodríguez, R. O. (2008). Logical approaches to fuzzy similarity-based reasoning: an overview. In *Preferences and similarities* (pp. 75-128). Springer Vienna.

Güvenir, H. A., Demiröz, G., & Ilter, N. (1998). Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals. *Artificial Intelligence in Medicine*, *13*(3), 147-165.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques*. Elsevier.

Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *160*(3), 523-541

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT press.

Harper, P. R. (2005). A review and comparison of classification algorithms for medical decision making. *Health Policy*, *71*(3), 315-331.

Hausdorff, F. (1914). Bemerkung über den Inhalt von Punktmengen. *Mathematische Annalen*, *75*(3), 428-433.

Hirota, K., & Pedrycz, W. (1991). Matching fuzzy quantities. *IEEE Transactions on Systems, Man, and Cybernetics*, *21*(6), 1580-1586.

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*, *37*(4), 543-558.

Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, *32*(10), 2513-2522.

Jacas, J. (1990). Similarity relations-the calculation of minimal generating families. *Fuzzy Sets and Systems*, *35*(2), 151-162.

Jacas, J., & Recasens, J. (1995). Fuzzy T-transitive relations: eigenvectors and generators. *Fuzzy Sets and Systems*, *72*(2), 147-154.

Janowicz, K., Raubal, M., & Kuhn, W. (2015). The semantics of similarity in geographic information retrieval. *Journal of Spatial Information Science, 2*, 29-57.

Jimenez, S., Becerra, C., & Gelbukh, A. (2012, June). Soft cardinality: A parameterized similarity function for text comparison. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation* (pp. 449-453). Association for Computational Linguistics.

Klawonn, F., & Castro Peña, J. L. (1995). Similarity in fuzzy reasoning. *Mathware & Soft Computing, 2*(3), 197-228.

Klawonn, F. (2003). Should fuzzy equality and similarity satisfy transitivity? Comments on the paper by M. De Cock and E. Kerre. *Fuzzy Sets and Systems, 133*(2), 175-180.

Klement, E. P., Mesiar, R., & Pap, E. (2000). *Triangular norms*. Springer Science & Business Media.

Larose, D. T. (2014). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons.

Le Capitaine, H., & Frelicot, C. (2009a). Classification with reject options in a logical framework: a fuzzy residual implication approach. In *International Fuzzy Systems Association World Congress and European Society for Fuzzy Logic and Technologies Conference* (pp. 855-860).

Le Capitaine, H., & Frelicot, C. (2009b). Towards a unified logical framework of fuzzy implications to compare fuzzy sets. In *International Fuzzy Systems Association World Congress and European Society for Fuzzy Logic and Technologies Conference* (pp. 1200-1205).

Le Capitaine, H. (2012). A relevance-based learning model of fuzzy similarity measures. *IEEE Transactions on Fuzzy Systems, 20*(1), 57-68.

Levner, I. (2005). Feature selection and nearest centroid classification for protein mass spectrometry. *BMC Bioinformatics, 6*(1), 1.

Li, S. T., & Ho, H. F. (2009). Predicting financial activity with evolutionary fuzzy case-based reasoning. *Expert Systems with Applications*, *36*(1), 411-422.

Li, P., & Jin, Q. (2012). Fuzzy relational equations with min-biimplication composition. *Fuzzy Optimization and Decision Making*, *11*(2), 227-240.

Lichman, M. (2013). UCI Machine Learning Repository. Link: http://archive.ics.uci.edu/ml/. [Pristupljeno 11.02.2016]

Lowen, R. (1996). Fuzzy set theory: Basic concepts, techniques and bibliography. Springer Science & Business Media.

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, *28*(5), 823-870.

Luukka, P. (2005). Similarity measure based classification. Doctoral thesis. Lappeenranta University of Technology, Lappeenranta.

Luukka, P., & Leppälampi, T. (2006). Similarity classifier with generalized mean applied to medical data. *Computers in Biology and Medicine*, *36*(9), 1026-1040.

Luukka, P. (2009). Classification based on fuzzy robust PCA algorithms and similarity classifier. *Expert Systems with Applications*, *36*(4), 7463-7468.

Luukka, P. (2011). Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications*, *38*(4), 4600-4607.

Luukka, P., Saastamoinen, K., & Kononen, V. (2001). A classifier based on the maximal fuzzy similarity in the generalized Lukasiewicz-structure. In *Fuzzy Systems, 2001. The 10th IEEE International Conference on* (Vol. 1, pp. 195-198). IEEE.

Luukka, P., & Kurama, O. (2013). Similarity classifier with ordered weighted averaging operators. *Expert Systems with Applications*, *40*(4), 995-1002.

Koloseni, D. (2015). *Differential evolution based classification with pool of distances and aggregation operators*. Doctoral thesis. Lappeenranta University of Technology, Lappeenranta.

Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, *31*, 249-268.

Klir, G., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic (Vol. 4)*. New Jersey: Prentice hall.

Kriegel, H. P., & Schubert, M. (2006, January). Advanced prototype machines: Exploring prototypes for classification. In *Proceedings of the 2006 SIAM International Conference on Data Mining* (pp. 176-187). SIAM.

Krishnapuram, R., Keller, J. M., & Ma, Y. (1993). Quantitative analysis of properties and spatial relations of fuzzy image regions. *IEEE Transactions on Fuzzy Systems*, *1*(3), 222-233.

McDermott, D. (1987). A critique of pure reason. *Computational Intelligence*, *3*(1), 151-160.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, No. 14, pp. 281-297).

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, *85*(3), 207.

Milligan, G. W. & Cooper, M. C. (1988). A study of standardization of variables in cluster analysis. *Journal of Classification 5* (2), 181–204.

Milošević, P., Petrović, B., Radojević, D., & Kovačević, D. (2014). A software tool for uncertainty modeling using Interpolative Boolean algebra. *Knowledge-Based Systems, 62*, 1-10.

Milošević, P., Poledica, A., Dragović, I., Petrović, B., & Radojević, D. (2013). Logic-based similarity measures for consensus. In *Proceedings of the XI Balkan conference on operational research BALCOR 2013* (pp. 473-481). University of Belgrade, Faculty of Organizational Sciences.

Minsky, M. (1975). A Framework for Representing Knowledge. In *The Psychology of Computer Vision* (pp. 211–277). McGraw-Hill.

Mitchell, M. T. (1997). *Machine learning*. McGraw-Hill.

Mizumoto, M. (1989). Pictorial representations of fuzzy connectives, part I: cases of t-norms, t-conorms and averaging operators. *Fuzzy Sets and Systems, 31*(2), 217-242.

Moore, A. W., & Lee, M. S. (1994, July). Efficient Algorithms for Minimizing Cross Validation Error. In *Proceedings of the Eighth International Conference on Machine Learning ICML 1994* (pp. 190-198). Morgan Kaufmann.

Moser, B. (2006). On representing and generating kernels by fuzzy equivalence relations. *The Journal of Machine Learning Research*, *7*, 2603-2620.

Niittymäki, J., & Turunen, E. (2003). Traffic signal control on similarity logic reasoning. *Fuzzy Sets and Systems*, *133*(1), 109-131.

Nguyen, T. T., & Armitage, G. (2008). A survey of techniques for internet traffic classification using machine learning. *IEEE Communications Surveys & Tutorials*, *10*(4), 56-76.

Petrović, B. J. (1998). *Teorija sistema*. Beograd: Fakultet organizacionih nauka.

Poincaré, H. (1902). *La Science et l'hypothese*. Paris: Flammarion.

Poincaré, H. (1958). *La valeur de la Science*, Flammarion, Paris (1904). English translation: *The Value of Science*.

Poledica, A., Bogojević-Arsić, V., & Petrović, B. (2010). Logical Aggregation as Similarity Measure in Case-based Reasoning. In *Computational Intelligence: Foundations and Applications - Proceedings of the 9th international FLINS conference*, (pp. 585-590).

Poledica, A., Milošević, P., Dragović, I., Petrović, B., & Radojević, D. (2015). Modeling consensus using logic-based similarity measures. *Soft Computing*, *19*(11), 3209–3219.

Poledica, A., Marković, D., & Živančević, S. (2016). Logical classification method for bankruptcy prediction. In *Proceedings of the XV International Symposium SymOrg 2016* (pp. 213-220). Belgrade: Faculty of Organizational Sciences.

Preparata, F. P., & Shamos, M. I. (1985). Introduction. In *Computational Geometry* (pp. 1-35). Springer New York.

Pyle, D. (1999). *Data preparation for data mining* (Vol. 1). Morgan Kaufmann.

Radojević, D. (2000). [0, 1]-valued logic: a natural generalization of Boolean logic. *Yugoslav Journal of Operations Research*, *10*(2), 185-216.

Radojević, D. & Petrović, B. (2001). Logical Aggregation of [0,1]-valued Logic. In *Proceedings of a Workshop on Computational Intelligence: Theory and Application* (pp. 69-80). Elektronski fakultet Niš.

Radojević, D. (2008a). Logical aggregation based on interpolative Boolean algebra, *Mathware & Soft Computing*, *15*, 125-141.

Radojević, D. (2008b). Real sets as consistent Boolean generalization of classical sets. In *From natural language to soft computing: New paradigms in artificial intelligence* (pp. 150-171). Editing House of Romanian Academy.

Radojević, D. (2010). Generalized (Real-Valued) order and Equivalence Relations. In *Proceedings of the 37th symposium on operational research SYM-OP-IS 2010* (pp 451-454). Media centar "Odbrana", Beograd.

Radojević, D. (2013). Real-valued realizations of Boolean algebra are a natural frame for consistent fuzzy logic. In *On Fuzziness: A Homage to Lotfi A. Zadeh – Volume 2. Studies in Fuzziness and Soft computing 299* (pp. 559-565). Springer-Verlag Berlin Heidelberg.

Radojević, D. (2015). Bulovski konzistenta NPN logika. In *Proceedings of the 42nd International Symposium on Operations Research SYM-OP-IS 2015* (pp. 354-357). Matematički Institut SANU, Beograd.

Rakićević, A. (2012). *Logičko klasterovanje - Hijerarhijski pristup*. Master teza. Univerzitet u Beogradu, Fakultet organizacionih nauka.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive psychology*, *3*(3), 382-407.

Richter, M. M., & Aamodt, A. (2005). Case-based reasoning foundations. *The Knowledge Engineering Review*, *20*(03), 203-207.

Rissland, E. L. (2006). AI and similarity. *IEEE Intelligent Systems*, *21*(3), 39-49.

Rorissa, A. (2007). Relationships between perceived features and similarity of images: A test of Tversky's contrast model. *Journal of the American Society for Information Science and Technology*, *58*(10), 1401-1418.

Rosch, E. (1983). Prototype classification and logical classification: The two systems. In *New trends in conceptual representation: Challenges to Piaget's theory* (pp. 73-86). Lawrence Erlbaum Associates Inc.

Saastamoinen, K. (2004a, December). On the Use of Generalized Mean with T-norms and T-conorms. In *Proceedings of 2004 IEEE Conference on Cybernetics and Intelligent Systems* (Vol. 1, pp. 569-573). IEEE.

Saastamoinen, K. (2004b, December). Semantic study of the use of parameterized S implications and equivalences in comparison. In In *Proceedings of 2004 IEEE Conference on Cybernetics and Intelligent Systems* (Vol. 1, pp. 574-577). IEEE.

Saastamoinen, K. (2006). *Many Valued Algebraic Structures as the Measures for Comparison*. IOS Press Amsterdam.

Saastamoinen, K., Luukka, P., & Könönen, V. (2002a). Weighted Fuzzy Similarity Classifier in the Łukasiewicz-Structure. In *Proceedings of the Third International Conference on Fuzzy Sets and Fuzzy Systems FSFS'02* (pp. 128-135).

Saastamoinen, K., Könönen, V., & Luukka, P. (2002b). A classifier based on the fuzzy similarity in the Łukasiewicz-structure with different metrics. In *Proceedings of the FUZZ-IEEE 2002 Conference* (pp. 363-367).

Saastamoinen K. and Sampo J. (2004). On General Class of Parameterized 3-Π Uninorm Based Comparison. *WSEAS Transactions on Mathematics, 3*(3), 482-486.

Santini, S. & Jain, R. (1999). Similarity Measures, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*(9), 871-883.

Simpson, P. K., 1992. Fuzzy min-max neural networks. I. Classification. *IEEE Transactions on Neural Networks, 3*(5), 776–786.

Shazmeen, S. F., Baig, M. M. A., & Pawar, M. R. (2013). Performance evaluation of different data mining classification algorithm and predictive analysis. *Journal of Computer Engineering*, *10*(6), 01-06.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317-1323.

Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., & Johannes, R. S. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the Annual Symposium on Computer Application in Medical Care* (pp. 261-265). American Medical Informatics Association.

Steinhaus, H. (1956). Sur la division des corp materiels en parties. *Bulletin de l'Academie Polonaise des Sciences*, *12*, 801-804.

Street, W. N., Wolberg, W. H., & Mangasarian, O. L. (1993). Nuclear feature extraction for breast tumor diagnosis. In *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology* (pp. 861-870). International Society for Optics and Photonics.

Sung, T. K., Chang, N., & Lee, G. (1999). Dynamics of modeling in data mining: interpretive approach to bankruptcy prediction. *Journal of Management Information Systems*, *16*(1), 63-85.

Teknomo, K. (2006). *Similarity measurement*. Dostupno na: http://people.revoledu.com/kardi/tutorial/Similarity/ [Pristupljeno 17.9.2015]

Tversky, A. & Krantz, D.H. (1970). The Dimensional Representation and the Metric Structure of Similarity Data. *Journal of Mathematical Psychology*, 572-596.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327-352.

Valverde, L. (1985). On the structure of F-indistinguishability operators. *Fuzzy Sets and Systems*, *17*(3), 313-328.

Veenman, C. J., & Reinders, M. J. (2005). The nearest subclass classifier: A compromise between the nearest mean and nearest neighbor classifier. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(9), 1417-1429.

Wang, X., De Baets, B., & Kerre, E. (1995). A comparative study of similarity measures. *Fuzzy Sets and Systems*, *73*(2), 259-268.

Wang, W. J. (1997). New similarity measures on fuzzy sets and on elements. *Fuzzy Sets and Systems*, *85*(3), 305-309.

Wolberg, W. H., & Mangasarian, O. L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proceedings of the National Academy of Sciences*, *87*(23), 9193-9196.

Zadeh, L.A. (1965). Fuzzy Sets, *Information and Control*, *8*, 338-353.

Zadeh, L. A. (1971). Similarity relations and fuzzy orderings. *Information Sciences*, *3*(2), 177-200.

Zadeh, L. A. (1977). Fuzzy Sets and Their Application to Pattern Classication And Clustering Analysis. *in J. Van Ryzin (Ed.): Classication and Clustering, Academic Press*, 251-299.

Zakon o računovodstvu (2013). Službeni glasnik Republike Srbije, br. 62/2013. Link: http://www.nbs.rs/. [Pristupljeno 10.04.2016.]

Zimmermann, H. J., & Zysno, P. (1980). Latent connectives in human decision making. *Fuzzy Sets and Systems*, *4*(1), 37-51.

Zimmermann, H. J. (2010). Fuzzy set theory. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(3), 317-332.

Zwick, R., Carlstein, E., & Budescu, D. V. (1987). Measures of similarity among fuzzy concepts: A comparative analysis. *International Journal of Approximate Reasoning*, *1*(2), 221-242.

12. СПИСАК СЛИКА

13. СПИСАК ТАБЕЛА

БИОГРАФИЈА

Ана Поледица је рођена 25. априла 1983. године у Чачку, Република Србија. Основну и средњу школу завршила је у Чачку као носилац дипломе „Вук Караџић“ и других специјалних диплома остварених на такмичењима из математике и других предмета.

Факултет организационих наука - Одсек за информационе системе и технологије уписала је 2002. године. Основне студије је завршила у року 2007. године са просечном оценом 9,79. Дипломирала је са оценом 10 на тему „Методолошки поступак процеса развоја веб апликација“. Носилац је бројних награда и признања за постигнут успех током студирања: добитник награде „Студент генерације“ на Факултету организационих наука у школској 2006/07. години од стране Универзитета у Београду, проглашена за најбољег дипломираног студента у школској 2006/07. години од стране Факултета организационих наука, стипендиста Министарства просвете и спорта, Фонда за младе таленте, општине Чачак, итд. Током студирања, од 2006. до средине 2007. године изводила је лабораторијске вежбе на предмету „Теорија система“.

У току 2007. године била је запослена у компанији „Soprex“ д.о.о. на позицији „Софтвер пројектант/програмер“. У компанији „S&T Srbija“ д.о.о. радила је наредних годину дана до краја 2008. године на позицији „Пословни (SAP) консултант“.

Од 18.11.2008. године запослена је на Факултету организационих наука, у звању Сарадника у настави, а затим у звању Асистента за ужу научну област Управљање системима. Током рада на факултету учествовала је у припреми и извођењу наставе (вежби/лабораторијских вежби и предавања) на основним и мастер студијама на предметима Теорија система, Фази логика и системи, Неуронске мреже и системи, Теорија система 2, Динамика организационих система и Моделовање и управљање организационим системима. Приликом евалуације од стране студената њен педагошки рад је оцењиван високим оценама о чему постоји писана евиденција на Факултету организационих наука. Током рада на Факултету стручно се усавршавала у оквиру пројекта „TRAIN“, у организацији Универзитета у Београду под покровитељством Фондације Краља Бодуена, са циљем унапређења научних, предавачких и педагошких вештина.

Школске 2008/2009. године уписала је докторске студије на Факултету организационих наука, студијски програм Управљање системима. Објавила је један научни рад из категорије М22 и више научних радова на међународним и домаћим скуповима и конференцијама. Награђена је од стране европске асоцијације за фази логику и технологију (EUSFLAT) за рад „A Consensus Model Based on Interpolative Boolean Algebra“ о чему је објављен истраживачки извештај у научном магазину „Mathware & Soft Computing Magazine“ у броју Vol. 21(1). Била је члан програмског одбора на светској конференцији посвећеној фази логици IFSA-EUSFLAT 2015. Члан је међународне асоцијације IEEE и европског друштва за фази логику и технологију – EUSFLAT.

Говори енглески језик и служи се немачким језиком.

Један је од оснивача хора ПолиФОН састављеног од запослених на Факултету организационих наука формираног 2014. године. Била је члан хора Обилић, АКУД „Бранко Крсмановић“.

ПРЕГЛЕД ОБЈАВЉЕНИХ РАДОВА

Ана Поледица је, у сарадњи са другим ауторима, објавила више научних радова у међународним часописима и зборницима радова са међународних конференција, а који су у директној вези са темом ове докторске дисертације.

Радови објављени у научним часописима међународног значаја (М22)

- **Poledica A.**, Milošević P., Dragović I., Petrović B., Radojević D.: *Modeling Consensus using Logic-based Similarity Measures*, - Soft Computing, 19(11), 2015, pp. 3209-3219. **DOI:**10.1007/s00500-014-1476-5. ISSN: 1432-7643 (Print), ISSN: 1433-7479 (Online), IF(2015)= 1.630

Радови саопштени на скупу међународог значаја

- **Poledica, A.**, Marković D., Živančević, S.: „*Logical classification method for bankruptcy prediction*", - Proceedings of the XV International Symposium - SymOrg 2016, Златибор, Србија, pp. 213-220, ISBN 978-86-7680-326-2.
- Milošević, P., **Poledica, A.**, Rakićević, A., Petrović, B., Radojević, D.: *Introducing Interpolative Boolean Algebra into Intuitionistic Fuzzy Sets*, - Proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology - IFSA-EUSFLAT-15, Gijon, Spain, DOI:10.2991/ifsa-eusflat-15.2015.1961389-1394. ISBN:978-94-62520-77-6 (online) ISSN: 1951-6851.
- **Poledica A**., Milošević P., Dragović I., Radojević D., Petrović B.: *A Consensus Model based on Interpolative Boolean Algebra*, - Proceedings of VIII International EUSFLAT Conference - EUSFLAT 2013, Milano, Italy, pp. 648-654. ISBN-978-90786-77-78-9.
- Milošević P., **Poledica A.**, Dragović I., Radojević D., Petrović B.: *Logic-based Similarity Measures for Consensus*, - Proceedings of XI Balkan Conference on Operational Research - BALCOR 2013, Belgrade and Zlatibor, Serbia, pp. 473-481. ISBN-978-86-7680-285-2.
- Kostić J., Bakajac M., Milošević P., **Poledica A.**: *Ranking of Banks based on Logical Aggregation*, - Proceedings of XI Balkan Conference on Operational Research - BALCOR 2013, Belgrade and Zlatibor, Serbia, pp. 2-11. ISBN-978-86-7680-285-2.
- Milošević P., Nešić I., **Poledica A.**, Radojević D., Petrović B.: *Models for Ranking Students: Selecting Applicants for Master of Science Studies*, -

Proceedings of V International Workshop Soft Computing Applications - SOFA 2012, Szeged, Hungary, pp. 93-103. ISBN:978-3-642-33940-0.

- **Poledica A.**, Rakićević A., Radojević D.: *Multi-Expert Decision Making Using Logical Aggregation*, - Proceedings of X International FLINS Conference - FLINS 2012, Istanbul, Turkey, pp. 561-566. ISBN: 978-981-4417-73-0.
- **Poledica A.**, Bogojević-Arsić V., Petrović B.: *Logical Aggregation as Similarity Measure in Case-based Reasoning*, - Proceedings of IX International FLINS Conference - FLINS 2010, Chengdu(EMei), Chine.
- Dobrić V., **Poledica A.**, Petrović B.: *Supply Chain Performance Measurement Using Logical Aggregation*, - Proceedings of IX International FLINS Conference - FLINS 2010, Chengdu(EMei), Chine.
- Medić K., Krstić S., **Poledica A.**: *Forex Market Exchange Rate Forecasting Using Technical Indicator Analysis with Neural Networks*, - Proceedings of IX Balkan Conference on Operational Research - BALCOR 2009, Constanta, Romania.

Радови објављени у научним часописима националног значаја

- Rakićević A., Milošević P., **Poledica A.**: *Sistem za evaluaciju finansijskih performansi kompanije zasnovan na logičkom pristupu*, - Journal of Information technology and multimedia systems Info M, 13(51), 2014, pp. 48-54. ISSN 1451-4397. *(M51)*
- Nešić, I., Rakićević, A., **Poledica, A.**, Petrović, B.: *Gaussian Variable Neighborhood Search and Enhanced Genetic Algorithm for Continuous Optimization*, - Electronic Notes in Discrete Mathematics, 39, 2012, pp. 273-280. ISBN: 1571-0653. *(M50)*

Радови саопштени на скупу националног значаја

- Dobrota M., Horvat A., **Poledica A.**: *Students' Comprehension of E-learning*, - zbornik radova radova XIX Međunarodne konferencije o informacionim i komunikacionim tehnologijama - YU INFO 2013, Kopaonik, Srbija.
- Kostić J., Petković T., **Poledica A.**, Radojević D.: *Swot Analysis Using AHP Approach With Interpolative Boolean Algebra*, - zbornik radova XXXIX Simpozijuma o operacionim istraživanjima - SYM-OP-IS 2012, Tara, Srbija, pp. 369-372, ISBN-978-86-7488-086-9.

- Dobrota M., **Poledica A.**, Bulajić M., Petrović B.: *Modeling Volatility Using Garch Model: Nasdaq-100 Application*, - zbornik radova XVIII Međunarodne konferencije o informacionim i komunikacionim tehnologijama - YU INFO 2012, Kopaonik, Srbija, pp. 18-23. ISBN:978-86-85525-09-4.
- Rakićević A., Nešić I., **Poledica A.**: *Forecasting Stock Performance Using Multi-Layer Feed-Forward Neural Network: Belgrade Stock Exchange Case*, - zbornik radova XIII Međunarodnog simpozijuma o organizacionim naukama - SYMORG 2012, Zlatibor, Srbija, pp. 802-808. ISBN: 978-86-7680-255-5.
- **Poledica, A.**, Dobrić, V., Kovačević, D.: *Predviđanje kretanja tržišnih indeksa korišćenjem veštačkih neuronskih mreža,* - zbornik radova XXXVIII Simpozijuma o operacionim istraživanjima - SYM-OP-IS 2011, Zlatibor, pp. 421-424, ISBN-978-86-403-1168-7.
- **Poledica A.**, Bogojević-Arsić V., Petrović B.: *Strateško upravljanje portfolio matricom i logičkom agregacijom*, - zbornik radova XXXVII Simpozijuma o operacionim istraživanjima - SYM-OP-IS 2010, Tara, Srbija, pp. 447-450, ISBN 978-86-335-0299-3.

Истраживачки извештаји

- **Poledica A.**, Milošević P., Dragović I., Radojević D., Petrović B.: „*A Consensus Model based on Interpolative Boolean Algebra*“, Mathware & Soft Computing Magazine Vol. 21, Issue 1, 2013, Scientific Report, (pp. 26-27. ISSN:1134-5632).

Изјава о ауторству

Име и презиме аутора Ана Поледица

Број индекса 35/2008

Изјављујем

да је докторска дисертација под насловом

Логички приступ моделовању сличности

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршила ауторска права и користила интелектуалну својину других лица.

Потпис аутора

У Београду, 8.7.2016.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Ана Поледица

Број индекса 35/2008

Студијски програм Управљање системима

Наслов рада Логички приступ моделовању сличности

Ментор проф. др Братислав Петровић

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предала ради похрањена у **Дигиталном репозиторијуму Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, 8.7.2016.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Логички приступ моделовању сличности

која је моје ауторско дело.

Дисертацију са свим прилозима предала сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучила.

1. Ауторство (CC BY)

2. Ауторство – некомерцијално (CC BY-NC)

3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)

4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)

5. Ауторство – без прерада (CC BY-ND)

6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци.
Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, 8.7.2016.

1. **Ауторство**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. **Ауторство – некомерцијално**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. **Ауторство – некомерцијално – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. **Ауторство – некомерцијално – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. **Ауторство – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. **Ауторство – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.