



UNIVERZITET U NOVOM SADU  
FAKULTET TEHNIČKIH NAUKA U  
NOVOM SADU



Sandra Nemet

---

**POVEĆANJE RAZUMLJIVOSTI MODELA  
ZASNOVANOG NA PRAVILIMA**

---

DOKTORSKA DISERTACIJA

NOVI SAD, 2020.





УНИВЕРЗИТЕТ У НОВОМ САДУ • ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА  
21000 НОВИ САД, Трг Доситеја Обрадовића 6

## КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Редни број, <b>РБР:</b>	
Идентификациони број, <b>ИБР:</b>	
Тип документације, <b>ТД:</b>	Монографска документација
Тип записа, <b>ТЗ:</b>	Текстуални штампани материјал
Врста рада, <b>ВР:</b>	Докторска дисертација
Аутор, <b>АУ:</b>	Сандра Немет
Ментор, <b>МН:</b>	Проф. др Стеван Станковски
Наслов рада, <b>НР:</b>	Повећање разумљивости модела заснованог на правилима
Језик публикације, <b>ЈП:</b>	Српски
Језик извода, <b>ЈИ:</b>	Српски/Енглески
Земља публиковања, <b>ЗП:</b>	Република Србија
Уже географско подручје, <b>УГП:</b>	АП Војводина, Нови Сад
Година, <b>ГО:</b>	2020
Издавач, <b>ИЗ:</b>	Ауторски репринт
Место и адреса, <b>МА:</b>	Факултет техничких наука, 21000 Нови Сад, Трг Доситеја Обрадовића 6
Физички опис рада, <b>ФО:</b> <small>(поглавља/страна/ цитата/табела/слика/графика/прилога)</small>	8/150/106/24/57/0/0
Научна област, <b>НО:</b>	Мехатроника
Научна дисциплина, <b>НД:</b>	Мехатроника, роботика и аутоматизација
Предметна одредница/Кључне речи, <b>ПО:</b>	Асоцијативна правила, Такаги-Сугено-Канг фази модел, Интерпретабилност, Анализа саобраћајних удеса, Визуелизација правила, Селекција карактеристика
<b>УДК</b>	
Чува се, <b>ЧУ:</b>	Библиотека Факултета техничких наука у Новом Саду
Важна напомена, <b>ВН:</b>	



## КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Извод, ИЗ:

Технике машинског учења имају бројне примене у техници, медицини, банкарству и многим другим областима. Присутан је пораст потребе за рачунарским методама који из велике количине података могу генерисати тачне и разумљиве моделе. Методе засноване на фази логици су добро познат приступ у моделовању сложених проблема. Такаги-Сугено-Канг (ТСК) фази модел заснован на правилима одликује висока тачност предикције, међутим његова разумљивост је слабија у односу на Мамдани фази модел.

Ова дисертација се бави истраживањем и пројектовањем радног оквира за побољшавање разумљивости ТСК фази модела. ТСК фази правила се алгоритамски комбинују са пронађеним асоцијативним правилима како би се побољшала целокупна разумљивост ТСК фази модела. Подударане ТСК фази правила са асоцијативним правилима се заснива на преклапању вредности атрибута асоцијативног правила и вредности одговарајућих улазно-излазних променљивих ТСК фази правила. У току процеса провере преклапања променљивих, променљиве из асоцијативних правила су представљене као фази синглтони. Све променљиве ТСК фази правила су апроксимирани Гаусовим фази функцијама припадности које су центриране у одговарајућим координатама центара кластера простора улазно-излазних варијабли. Поступку агрегације ТСК фази правила и асоцијативних правила претходи фаза селекције обележја. У ту сврху је урађена стохастичка претрага мулти-димензионалног простора улазно-излазних варијабли проблема комбиновањем метода оптимизације јата честица и вишеслојног перцептрона.

Увођењем нових компоненти поред ТСК фази модела, развијени радни оквир омогућава бољу разумљивост ТСК фази правила. Такође, он обезбеђује јаснију идентификацију и разумевање значајних карактеристика моделованог система или процеса, као и скривених релација између његових улазних обележја и последица. Поред једноставног увида у везе међу правилима, радни оквир обезбеђује процену важности фази правила у односу према осталим правилима и према комплетном ТСК фази моделу. Предложени радни оквир је развијен и илустрован на проблему анализе саобраћајних незгода са учешћем пешака. Помоћу њега је омогућено подробније разумевање развијеног модела заснованог на правилима, као и јаснија идентификација значајних узрока саобраћајних незгода.

Датум прихватања теме, ДП:

Датум одбране, ДО:

Чланови комисије,  
КО:

Председник: Др Гордана Остојић, редовни професор

Члан: Др Љубомир Миладиновић, редовни професор

Члан: Др Драган Јовановић, редовни професор

Члан: Др Ласло Тарјан, доцент

Члан: Др Срђан Тегелтија, доцент

Члан, ментор: Др Стеван Станковски, редовни професор

Потпис ментора



UNIVERSITY OF NOVI SAD • FACULTY OF TECHNICAL SCIENCES  
21000 NOVI SAD, Trg Dositeja Obradovića 6

## KEY WORDS DOCUMENTATION

Accession number, <b>ANO</b> :	
Identification number, <b>INO</b> :	
Document type, <b>DT</b> :	Monographic publication
Type of record, <b>TR</b> :	Textual printed material
Contents code, <b>CC</b> :	PhD thesis
Author, <b>AU</b> :	Sandra Nemet, Msc.
Mentor, <b>MN</b> :	Prof. Stevan Stankovski, PhD
Title, <b>TI</b> :	Increasing the understandability of a rule-based model
Language of text, <b>LT</b> :	Serbian
Language of abstract, <b>LA</b> :	Serbian/English
Country of publication, <b>CP</b> :	Republic of Serbia
Locality of publication, <b>LP</b> :	AP Vojvodina, Novi Sad
Publication year, <b>PY</b> :	2020
Publisher, <b>PB</b> :	Author's reprint
Publication place, <b>PP</b> :	Faculty of Technical Sciences, 21000 Novi Sad, Trg Dositeja Obradovića 6
Physical description, <b>PD</b> : <small>(chapters/pages/ref./tables/pictures/graphs/appendix)</small>	8/150/106/24/57/0/0
Scientific field, <b>SF</b> :	Mechatronics
Scientific discipline, <b>SD</b> :	Mechatronics, Robotics and Automation
Subject/Key words, <b>S/KW</b> :	Association rules, Takagi-Sugeno-Kang fuzzy model, Interpretability, Traffic accident analysis, Rules visualization, Feature selection
<b>UC</b>	
Holding data, <b>HD</b> :	Library of the Faculty of Technical Sciences, University of Novi Sad
Note, <b>N</b> :	



**KEY WORDS DOCUMENTATION**

**Abstract, AB:**

The amount and diversity of machine learning applications have numerous and diverse applications in different domains, including technology, medicine, finance and so on. There is an enhanced need for computational methods which can generate accurate and interpretable models from big data. Fuzzy logic methods are a well-known approach in modelling of complex problems. Takagi-Sugeno-Kang (TSK) rule-based fuzzy model is characterized by high prediction accuracy, however its interpretability is lesser in comparison to the Mamdani fuzzy model.

This PhD thesis deals with the framework design of an easily interpretable TSK fuzzy model. TSK fuzzy rules are algorithmically combined with related association rules in order to improve the overall interpretability of the TSK model. Matching TSK fuzzy rules with association rules is based on attribute values' overlapping of the association rule and corresponding input-output variables' value of the TSK fuzzy rule. During the process of variables' overlapping, variables from association rules are represented as fuzzy singletons. All TSK fuzzy rules variables are approximated with Gaussian fuzzy membership functions, which are centered in corresponding coordinated of cluster centers in the input-output variables' space. Aggregation of TSK fuzzy and association rules is preceded by features selection phase. Feature selection is performed using a stochastic search procedure of multi-dimensional input-output variables' space. The stochastic search method represents a hybrid method that combines particle swarm optimization and multi-layer perceptron.

By adding new components besides the TSK fuzzy model, the developed framework enables better interpretability of TSK fuzzy rules. It provides a more readable identification and understanding of significant characteristics of modeled system or process, as well as hidden relations between its input features and consequences. Besides a simple insight into the rules' interconnections of the two rule-based models, the framework provides an assessment of fuzzy rules' importance in both, and in accordance to other rules and to the complete TSK fuzzy model. The proposed framework is developed and illustrated on the problem of analysis of traffic accidents with pedestrian involvement. It provides a deeper understanding of the built rule-based model, as well as a more readable identification of significant accident causes.

Accepted by the Scientific Board on, **ASB:**

Defended on, **DE:**

Defended Board, **DB:**

President: Gordana Ostojić, PhD, Full professor

Member: Ljubomir Miladinović, PhD, Full professor

Member: Dragan Jovanović, PhD, Full professor

Member: Laslo Tarjan, PhD, Docent

Member: Srđan Tegeltija, PhD, Docent

Member, Stevan Stankovski, PhD, Full professor

Menthor's sign

## Zahvalnica

Zahvaljujem se svom mentoru, profesoru dr Stevanu Stankovskom, na usmeravanju, korisnim sugestijama i poverenju tokom izrade doktorske disertacije. Takođe, zahvaljujem se profesoricu dr Gordani Ostojić na pomoći i razumevanju, kao i profesoru dr Draganu Jovanoviću na saradnji početkom ovog naučnog istraživanja.

Veliku zahvalnost na moralnoj podršci dugujem mojim roditeljima tokom čitavih mojih studija, kao i suprugu koji mi je uvek davao dobar savet u prelomnim trenucima izrade disertacije. Naposljetku bih volela da posvetim ovu doktorsku disertaciju mojoj ćerkici Emiliji.





## Sažetak

Tehnike mašinskog učenja imaju brojne primene u tehnici, medicini, bankarstvu, ali i mnogim drugim oblastima. Prisutan je porast potrebe za računarskim metodama koji iz velike količine podataka mogu generisati tačne i razumljive modele. Metode zasnovane na fazi logici predstavljaju dobro poznat pristup u modelovanju složenih problema. Takagi-Sugeno-Kang (TSK) fazi model zasnovan na pravilima odlikuje visoka tačnost predikcije, međutim njegova razumljivost je slabija u odnosu na Mamdani fazi model.

Ova disertacija se bavi istraživanjem i projektovanjem radnog okvira za poboljšavanje razumljivosti TSK fazi modela. TSK fazi pravila se algoritamski kombinuju sa pronađenim asocijativnim pravilima kako bi se poboljšala celokupna razumljivost TSK fazi modela. Podudaranje TSK fazi pravila sa asocijativnim pravilima se zasniva na preklapanju vrednosti atributa asocijativnog pravila i vrednosti odgovarajućih ulazno-izlaznih promenljivih TSK fazi pravila. U toku procesa provere preklapanja promenljivih, promenljive iz asocijativnih pravila su predstavljene kao fazi singltoni. Sve promenljive TSK fazi pravila su aproksimirane Gausovim fazi funkcijama pripadnosti koje su centrirane u odgovarajućim koordinatama centara klastera prostora ulazno-izlaznih varijabli. Postupku agregacije TSK fazi pravila i asocijativnih pravila prethodi faza selekcije obeležja. U tu svrhu je urađena stohastička pretraga multi-dimenzionalnog prostora ulazno-izlaznih varijabli problema kombinovanjem metoda optimizacije jata čestica i višeslojnog perceptrona.

Uvođenjem novih komponenti, pored TSK fazi modela, razvijeni radni okvir omogućava bolju razumljivost TSK fazi pravila. Takođe, on obezbeđuje jasniju identifikaciju i razumevanje značajnih karakteristika modelovanog sistema ili procesa, kao i skrivenih relacija između njegovih ulaznih obeležja i posledica. Pored jednostavnog uvida u veze među pravilima, radni okvir obezbeđuje procenu važnosti fazi pravila u odnosu prema ostalim pravilima, kao i prema kompletnom TSK fazi modelu. Predloženi radni okvir je razvijen i ilustrovan na problemu analize saobraćajnih nezgoda sa učešćem pešaka. Pomoću njega je omogućeno detaljnije razumevanje razvijenog modela zasnovanog na pravilima, kao i jasnija identifikacija značajnih uzroka saobraćajnih nezgoda. Na osnovu identifikovanih faktora koji utiču na pojavu

saobraćajnih udesa, kao i njihovih međusobnih relacija, moguće je uvođenje pouzdanijih mera predostrožnosti koje će uticati na smanjenje broja i težine posledica potencijalnih saobraćajnih nezgoda u budućnosti. Razvijeni radni okvir je moguće koristiti u mnogim oblastima modelovanja, analize modela i procesa odlučivanja, gde je razumevanje računarskih modela od velike važnosti.

## Abstract

The amount and diversity of machine learning applications have numerous and diverse applications in different domains, including technology, medicine, finance and so on. There is an enhanced need for computational methods which can generate accurate and interpretable models from big data. Fuzzy logic methods are a well-known approach in modeling complex problems. The Takagi-Sugeno-Kang (TSK) rule-based fuzzy model is characterized by high prediction accuracy, however its interpretability is lesser in comparison to the Mamdani fuzzy model.

This PhD thesis deals with the framework design of an easily interpretable TSK fuzzy model. TSK fuzzy rules are algorithmically combined with the related association rules in order to improve the overall interpretability of the TSK model. Matching of TSK fuzzy rules with association rules is based on attribute values' overlapping of the association rule and the corresponding input-output variables' value of the TSK fuzzy rule. During the process of variables' overlapping, variables from association rules are represented as fuzzy singletons. All TSK fuzzy rule variables are approximated with Gaussian fuzzy membership functions, which are centered in corresponding coordinated of cluster centers in the input-output variables' space. Aggregation of TSK fuzzy and association rules is preceded by features selection phase. Feature selection is performed using a stochastic search procedure of a multi-dimensional input-output variables' space. The stochastic search method represents a hybrid method that combines particle swarm optimization and the multi-layer perceptron.

By adding new components besides the TSK fuzzy model, the developed framework enables better interpretability of TSK fuzzy rules. It provides a more readable identification and understanding of significant characteristics of the modeled system or process, as well as hidden relations between its input features and consequences. Besides a simple insight into the rules' interconnections of the two rule-based models, the framework provides an assessment of fuzzy rules' importance in both, and in accordance to other rules and to the complete TSK fuzzy model. The proposed framework is developed and illustrated on the problem of analysis of traffic accidents with pedestrian involvement. It provides a deeper understanding of the built rule-based model, as well as a more readable identification of significant accident

causes. Based on the identified factors that influence the appearance of traffic accidents, as well as their inter-relations, more reliable countermeasures can be introduced that will have an impact on reducing the number of traffic accidents and the severity of their consequences in the future. The developed framework can be used in many domains of analysis modeling and decision-making processes where computational model interpretability is crucial.

## Sadržaj

1.	Uvod.....	1
1.1	Predmet istraživanja .....	1
1.2	Cilj istraživanja i hipoteze.....	4
1.3	Mogućnosti primene .....	6
1.4	Kratak sadržaj disertacije .....	8
2.	Pregled aktuelnog stanja u oblasti .....	11
2.1	Pregled istraživanja u poboljšavanju interpretabilnosti modela .....	11
2.2	Pregled istraživanja u projektovanju modela za predikciju saobraćajnih nesreća .....	13
2.2.1	Modeli zasnovani na veštačkim neuronskim mrežama .....	14
2.2.2	Modeli zasnovani na SVM .....	16
2.2.3	Modeli zasnovani na pravilima i fazi logici .....	16
2.2.4	Modeli zasnovani na asocijativnim pravilima.....	18
2.2.5	Hibridni modeli i komparacije više modela .....	19
3.	Pregled korišćenih algoritama .....	23
3.1	Veštačke neuronske mreže (VNM) .....	23
3.1.1	Neuronske mreže sa nadzorom – Višeslojni perceptroni.....	25
3.2	Samo-organizujuće neuronske mreže.....	28
3.2.1	Kohonen samo-organizujuće neuronske mreže .....	29
3.3	Algoritmi grupisanja .....	30
3.3.1	<i>K-means</i> algoritam klaster analize.....	31
3.4	Korelacija .....	32
3.5	Metoda optimizacije sa rojem čestica .....	33
3.6	Asocijativna pravila .....	36
3.6.1	Apriori algoritam.....	37

3.7	Fazi logika .....	40
3.8	Takagi-Sugeno-Kang fazi-logički modeli .....	42
3.8.1	Takagi-Sugeno-Kang fazi-logički model .....	43
3.8.2	Generisanje TSK fazi modela.....	44
3.8.3	Opis algoritma za generisanje TSK fazi logičkog modela .....	48
3.9	Vizualizacija pomoću Fingrams pristupa .....	48
4.	Predloženi radni okvir metodologije.....	53
4.1	Predobrada podataka .....	54
4.2	Asocijativna pravila i Apriori algoritam .....	56
4.3	TSK fazi model .....	56
4.4	Povezivač TSK asocijativnih pravila .....	58
4.5	Detaljan opis algoritma Povezivača TSK fazi pravila i asocijativnih pravila 61	
4.5.1	Algoritam interpretiranja TSK fazi modela sa asocijativnim pravilima 63	
4.5.2	Algoritam interpretiranja asocijativnih pravila sa TSK fazi pravilom .64	
4.5.3	Uprošćeni postupak određivanja sličnosti između asocijativnih i TSK fazi pravila 65	
5.	Rezultati selekcije obeležja .....	67
5.1	Opis baze podataka .....	67
5.2	Selekcija obeležja kros-korelacijom .....	74
5.3	Rezultati selekcije pomoću PSO+ANN algoritma .....	80
6.	Rezultati metodologije.....	89
6.1	Generisanje asocijativnih pravila Apriori algoritmom.....	89
6.2	Prvi TSK fazi model .....	92
6.3	Drugi TSK fazi model.....	112
6.4	Diskusija rezultata za oba TSK fazi modela.....	129
6.5	Rezultati vizualizacije - Fingrams.....	130

6.6	Komparativna analiza modela interpretabilnosti zasnovanih na pravilima	133
7.	Zaključak .....	137
8.	Literatura .....	141





## Spisak slika

Slika 3.1 Izgled višeslojne VNM.....	24
Slika 3.2 Primer VNM sa dva skrivena sloja .....	27
Slika 3.3 Sigmoidne aktivacione funkcije .....	28
Slika 3.4 Primer Kohonenove samo-organizujuće mreže .....	30
Slika 3.5 Struktura fazi sistema zaključivanja .....	40
Slika 3.6 Primer Fingrams grafa .....	51
Slika 4.1 Shema postupka poboljšanja interpretabilnosti modela .....	53
Slika 4.2 Ilustracija preklapanja promenljivih iz skupa fazi promenljivih FV i skupa promenljivih S .....	59
Slika 4.3 Komponente algoritma za povezivanje TSK fazi pravila i asocijativnih pravila ..	63
Slika 5.1 Matrica kros-korelacija za posmatranih 25 varijabli .....	76
Slika 5.2 Kros-korelacije za skupove podataka sa vrednostima “lake povrede” i “bez povreda” .....	77
Slika 5.3 Kros-korelacije za skupove podataka sa vrednostima “teške povrede” i “bez povreda” .....	79
Slika 5.4 Izbor pet varijabli sa PSO+ANN algoritmom.....	82
Slika 5.5 Konvergencija PSO+ANN algoritma prilikom izbora pet varijabli .....	82
Slika 5.6 Izbor četiri varijable sa PSO+ANN algoritmom.....	83
Slika 5.7 Izbor tri varijable sa PSO+ANN algoritmom .....	84
Slika 5.8 Izbor dve varijable sa PSO+ANN algoritmom .....	84
Slika 5.9 Pozicije rešenja podskupova selektovanih varijabli tokom rada PSO+ANN algoritma pri izboru 5 varijabli.....	85
Slika 5.10 Zastupljenost varijabli za selektovani interval od 2 do 7 varijabli .....	86
Slika 5.11 Ilustracija konvergencije PSO+ANN algoritma prilikom selekcije obeležja .....	87
Slika 6.1 Kumulativna sličnost asocijativnih pravila i prvog TSK fazi modela .....	94
Slika 6.2 Intezitet preklapanja posledičnih delova parova asocijativnih i TSK pravila .....	95
Slika 6.3 Preklapanja uslovnih delova svakog asocijativnog pravila sa svim pravilima prvog TSK fazi modela.....	96
Slika 6.4 Sličnost asocijativnih pravila sa odgovarajućim pravilom prvog TSK fazi modela .....	96

Slika 6.5 Presek uslovnih varijabli Pol, Prelaženje kolovoza i Starosne grupe.....	98
Slika 6.6 Presek uslovnih varijabli Pol, Prelaženje kolovoza i Starosne grupe.....	99
Slika 6.7 Presek uslovnih varijabli Pol, Prelaženje kolovoza i Starosne grupe.....	100
Slika 6.8 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza.....	101
Slika 6.9 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza.....	102
Slika 6.10 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza.....	103
Slika 6.11 Presek uslovnih varijabli Ponašanje i Starosne grupe .....	104
Slika 6.12 Presek uslovnih varijabli Ponašanje i Starosne grupe .....	105
Slika 6.13 Presek uslovnih varijabli Ponašanje i Starosne grupe .....	106
Slika 6.14 Presek uslovnih varijabli Pol, Pešački prelaz i Prelaženje kolovoza.....	107
Slika 6.15 Presek uslovnih varijabli Pol, Pešački prelaz i Prelaženje kolovoza.....	108
Slika 6.16 Presek uslovnih varijabli Pol, Pešački prelaz i Prelaženje kolovoza.....	109
Slika 6.17 Ilustracija postupka generisanja funkcija pokrivenosti TSK fazi pravila sa asocijativnim pravilima kod prvog TSK fazi modela.....	110
Slika 6.18 Pokrivenost pravila prvog TSK fazi modela sa asocijativnim pravilima .....	110
Slika 6.19 Kumulativna sličnost asocijativnih pravila i drugog TSK fazi modela .....	114
Slika 6.20 Intezitet preklapanja posledičnih delova asocijativnih i TSK fazi pravila .....	115
Slika 6.21 Nivo preklapanja uslovnih delova asocijativnih pravila sa uslovnim delovima TSK fazi pravila .....	115
Slika 6.22 Sličnost asocijativnih pravila sa odgovarajućim pravilom drugog TSK fazi modela .....	116
Slika 6.23 Presek uslovnih varijabli Pešački prelaz i Ponašanje .....	117
Slika 6.24 Presek uslovnih varijabli Pešački prelaz i Ponašanje .....	118
Slika 6.25 Presek uslovnih varijabli Pešački prelaz i Ponašanje .....	119
Slika 6.26 Presek uslovnih varijabli Pešački prelaz i Ponašanje .....	120
Slika 6.27 Presek uslovnih varijabli Pešački prelaz i Ponašanje .....	121
Slika 6.28 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza.....	122
Slika 6.29 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza.....	123
Slika 6.30 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza.....	124
Slika 6.31 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza.....	125
Slika 6.32 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza.....	126
Slika 6.33 Ilustracija postupka generisanja funkcija pokrivenosti fazi pravila sa asocijativnim pravilima kod drugog TSK fazi modela .....	127
Slika 6.34 Pokrivenost pravila drugog TSK fazi modela sa asocijativnim pravilima.....	127
Slika 6.35 <i>Fingrams</i> graf <i>GUAJE WM</i> fazi modela.....	130

Slika 6.36 <i>Fingrams</i> graf modela asocijativnih pravila .....	132
Slika 6.37 Vrednosti Naukov-og indeksa razmatranih modela zasnovanih na pravilima .....	135



## Spisak tabela

Tabela 3.1 Pseudo-kod osnovne verzije PSO algoritma .....	35
Tabela 3.2 Pseudo-kod osnovnog Apriori algoritma .....	38
Tabela 3.3 Ilustrativni primer Apriori postupka .....	39
Tabela 3.4 Opis algoritma obuke TSK fazi modela .....	48
Tabela 4.1 Algoritam povezivača TSK fazi i asocijativnih pravila .....	62
Tabela 4.2 Algoritam interpretiranja TSK fazi modela sa asocijativnim pravilima .....	64
Tabela 4.3 Algoritam interpretiranja asocijativnih pravila sa TSK fazi pravilom .....	65
Tabela 4.4 Algoritam određivanja sličnosti između asocijativnih i TSK fazi pravila .....	65
Tabela 5.1 Objašnjenja za grupu promenljivih: Vremenske karakteristike .....	68
Tabela 5.2 Objašnjenja za grupu promenljivih: Karakteristike udesa .....	69
Tabela 5.3 Objašnjenja za grupu promenljivih: Karakteristike puta .....	70
Tabela 5.4 Objašnjenja za grupu promenljivih: Karakteristike učesnika (personalni podaci) .....	72
Tabela 5.5 Polazne varijable filtriranog skupa podataka o saobraćajnim udesima sa učešćem pešaka .....	74
Tabela 5.6 Lista selektovanih parova varijabli sa najvećim promenama kros-korelacija: slučajevi sa lakim povredama i bez povreda .....	78
Tabela 5.7 Lista selektovanih parova varijabli sa najvećim promenama kros-korelacija: skupovi uzoraka sa vrednostima varijable Posledice: "teške povrede" i "bez povreda" .....	79
Tabela 6.1 Odabrana asocijativna pravila .....	90
Tabela 6.2 Definicija funkcija pripadnosti normalizovanog prostora razmatranja varijabli asocijativnih pravila .....	91
Tabela 6.3 Posledični deo parametara .....	92
Tabela 6.4 Koordinate centroida pravila .....	92
Tabela 6.5 Preklapanja asocijativnih pravila sa pravilima prvog TSK fazi modela .....	93
Tabela 6.6 Posledični deo parametara .....	112
Tabela 6.7 Koordinate centroida TSK fazi pravila .....	112
Tabela 6.8 Preklapanja asocijativnih pravila sa pravilima drugog TSK fazi modela .....	113
Tabela 6.9 Izračunate kvantitativne mere interpretabilnosti modela zasnovanih na pravilima .....	135



# 1. Uvod

## 1.1 Predmet istraživanja

Predmet istraživanja ove disertacije je pronalaženje načina za poboljšanje razumljivosti ponašanja fazi logičkih modela. Naime, fazi logički model može odlikovati zadovoljavajuća tačnost, ali i nedovoljna razumljivost. U tu svrhu je potrebno pronaći metodu za poboljšanje razumljivosti ponašanja modela (eng. *interpretability*). Odnos između tačnosti i razumljivosti modela je profesor L. Zadeh [1] definisao pomoću principa nekompatibilnosti realnog problema, i to na sledeći način: što je model tačniji, njegovo rešenje je nejasnije. Naime, povećavanjem tačnosti modela se povećava i njegova složenost. Porastom složenosti modela raste broj relacija i smanjuje se razumljivost ponašanja modela. Dakle, tačnost i razumljivost modela su dve kontradiktorne osobine i iznad određene granice postaju gotovo isključive. Razumljivost ponašanja modela se sveobuhvatnije definiše pojmom interpretabilnosti.

Sa interpretabilnošću se na razumljiv način izražava ponašanje realnog sistema i njegovog modela. Interpretabilnost je subjektivno svojstvo koje zavisi od procenitelja, a u vezi je sa nekoliko faktora. To su pre svega struktura modela, broj ulaznih promenljivih, broj pravila, njihov tip i slično. Bodenhofer i Bauer [2] su interpretabilnost definisali kao mogućnost procene ponašanja sistema samo na osnovu čitanja i razumevanja baze pravila. Međutim, još uvek ne postoji standardna i sveobuhvatna mera za procenu interpretabilnosti. Razlog tome je što nije moguće u potpunosti istovremeno optimizovati i tačnost i interpretabilnost, a neophodno je da obe karakteristike budu u određenoj meri zadovoljene [3, 4].

Za model zasnovan na fazi pravilima takođe važi princip nekompatibilnosti: što je model složeniji, to je njegova interpretabilnost manja. Cilj fazi modela je naravno, ne samo povećanje razumljivosti ponašanja, već pre svega težnja ka većoj tačnosti. Pronalazak optimalnog balansa između tačnosti i razumljivosti ponašanja predstavlja jedan od najtežih zadataka u oblasti projektovanja fazi logičkih modela [5].

Lipton [6] je definisao pojam interpretabilnosti modela sa konceptima koji spadaju u dve šire kategorije: transparentne modele (suprotno od modela crne

kutije) i modele naknadne (eng. *post-hoc*) interpretabilnosti. Model naknadne interpretabilnosti uključuje, pored osnovnog prediktivnog modela, i dopunski model koji objašnjava primarni model (eng. *explanatory model*). Takav dopunski model je odvojen od osnovnog prediktivnog modela i ima zadatak da transformiše podatke i/ili parametre polaznog opisnog modela, a po potrebi i da obavlja vizualizaciju modela. Pri tome je potrebno obratiti pažnju na to da se za objašnjavanje prediktivnog modela na bazi modela crne kutije ne treba koristiti dopunski model zasnovan takođe na modelu crne kutije [7]. Pored toga, transparentne prediktivne modele odlikuju tri moguće karakteristike: razgradivost, algoritamska transparentnost i mogućnost simuliranja [6]. Razgradivost se odnosi na viđenje i razumevanje delova modela i kako oni doprinose rezultatima prediktivnog modela. Mogućnost simuliranja se odnosi na modele koje je moguće mentalno simulirati ili zamisliti u razumnom vremenu.

Istraživački rad obuhvaćen ovom doktorskom disertacijom je prevashodno fokusiran na interpretiranje neodređenosti i nejasnih interakcija između ulaznih i izlaznih promenljivih fazi logičkih modela generisanih tehnikama mašinskog učenja. Naime, fazi modele, iako su zasnovani na logičkim pravilima, takođe može karakterisati prisustvo nejasnih relacija. Razlikuju se dva osnovna tipa modela zasnovanih na fazi logičkim pravilima. Kod lingvističkog fazi modelovanja (npr. Mamdani fazi model), eksperti stručno znanje direktno unose u fazi logički model i fokus je tada na razumljivosti ponašanja sistema. Pri tome, u toku postupka generisanja baze pravila postoje poteškoće usled složenosti problema ili nedostatka stručnjaka za datu problematiku [8, 9]. Sa druge strane, kod preciznog fazi modelovanja (npr. Takagi-Sugeno-Kang fazi model), gde se znanje identifikuje i izdvaja iz eksperimentalnih podataka, fokus je na tačnosti modela. Ovi tipovi fazi modela nisu međusobno isključivi, a hibridni pristupi su poznati i detaljno proučeni [10]. Na osnovu korišćenih podataka, generisanih fazi skupova i fazi pravila se može uraditi analiza i interpretiranje modela. Rezultat ovog postupka može biti zaključak da neka fazi pravila treba modifikovati ili ukloniti, odnosno da je potrebno izmeniti balans između tačnosti i razumljivosti ponašanja modela [11].

Celokupni postupak istraživanja je zasnovan na tehnici analize podataka, postupku koji se odvija u nekoliko faza koje se mogu ciklično ponavljati. Faze istraživanja su: (1) Razumevanje problema i podataka, (2) Priprema podataka, (3) Modelovanje i (4) Evaluacija.



Prva faza istraživanja obuhvata razumevanje problema, prikupljanje podataka i njihovo proučavanje. U okviru ove faze istraživanja su identifikovani interesantni postupci za analizu i generisanje modela. Pored toga, urađen je pregled postojećih postupaka i sumiran je pregled literature dosadašnjih istraživanja u oblastima od značaja: razumnjivosti ponašanja modela, modela za analizu i predikciju saobraćajnih nezgoda ili bezbednosti učesnika u saobraćaju.

Druga faza istraživanja obuhvata analizu podataka dobijenih iz baze podataka sa realnim informacijama o saobraćajnim nezgodama. Primer baze podataka definisan je saobraćajnim nezgodama u kojima su učestvovali pešaci u periodu od 2008. do 2011. godine u Novom Sadu. Ovaj inicijalni skup podataka sadrži veliki broj atributa sa potencijalnim uticajem na težinu saobraćajnih nezgoda u kojima su učestvovali pešaci. Ovaj skup podataka je prvo očišćen od suvišnih podataka koji su ili nepotpuni, redundantni ili irelevantni usled nedostajućih, nepotrebnih ili netačnih vrednosti. Nakon procesa filtriranja podataka sledi faza odabira ključnih karakteristika, tj. obeležja, koja će biti uključena u finalni podskup obeležja, spreman za dalju obradu.

U trećoj fazi, fazi modelovanja, su primenjene izabrane tehnike modelovanja i procene parametara modela. Potom je razvijena metodologija za povećanje razumljivosti ponašanja modela. Razvijena metodologija je zasnovana na tehnikama mašinskog učenja, poput neuronskih mreža, asocijativnih pravila i fazi logičkih modela. Postupak za povećavanje razumljivosti modela zasnovanog na pravilima predstavlja ključni deo faze modelovanja. Celokupan postupak se ukratko može opisati na sledeći način: U fazi predobrade se inicijalni podaci filtriraju i najuticajnije promenljive se selektuju koristeći pretragu prostora ulaznih promenljivih pomoću optimizacije sa rojem čestica (eng. *particle swarm optimization*) koja koristi višeslojni perceptron kao evaluator kvaliteta jedinki - rešenja. Koriste se podaci sa najuticajnijim karakteristikama kako bi se formirala dva modela zasnovana na pravilima. Prvi model koristi Takagi-Sugeno-Kang (TSK) fazi pravila, dok je drugi model zasnovan na asocijativnim pravilima. Generator asocijativnih pravila identifikuje najinteresantnije veze između odabranih ulaznih atributa i težine povreda pešaka, kao posledice nezgode koje figurišu u posledičnom delu svakog pravila. TSK fazi model generiše skup fazi pravila sa težinama povreda pešaka kao posledicama koristeći potpuno iste podatke. U cilju poboljšanja razumljivosti ponašanja TSK fazi pravila, uveden je novi algoritam koji predstavlja povezivač TSK fazi modela i asocijativnih pravila. Ulazi algoritma za povezivanje TSK fazi i asocijativnih pravila su parametri TSK fazi modela i skupa

asocijativnih pravila generisanih pomoću Apriori algoritma. Povezivač TSK fazi pravila i asocijativnih pravila se sastoji od dve komponente: Modula za pronalaženje povezanih TSK fazi i asocijativnih pravila koji uključuje proračun preklapanja vrednosti svih promenljivih koje figurišu u razmatranim parovima TSK fazi i asocijativnih pravila; i Interpreter TSK fazi i asocijativnih pravila koji predstavlja povezana pravila na adekvatan način.

U poslednjoj fazi istraživanja je urađena procena performansi kreiranih modela. Performanse modela su analizirane u svetlu njihove razumljivosti i tu su korišćene standardne mere interpretabilnosti, dok su za ocenu tačnosti korišćene mere kao što su korelacija, srednja ili apsolutna kvadratna greška. Dobijeni modeli i izvedeni zaključci omogućavaju da se proširi znanje o saobraćajnim nezgodama na osnovu ulaznih podataka i otkrivenih uzročno-posledičnih veza između obeležja i posledica.

## **1.2 Cilj istraživanja i hipoteze**

Cilj ovog istraživanja je da se razvije radni okvir kojim bi se poboljšala razumljivost ponašanja tačnog TSK fazi modela. Čitljivost TSK fazi pravila nije toliko visoka kao kod Mamdani pravila, uprkos njihovoj AKO-ONDA formi. Za to postoje bar dva razloga vezana za uzročni i posledični deo pravila. Uzročni deo TSK fazi pravila se sastoji od koordinata centra klastera podataka kojem je dato pravilo dodeljeno. Veze ulaznih varijabli sa odgovarajućim koordinatama centra klastera nisu očigledne jer figurišu u proračunima funkcije udaljenosti uzoraka i centara klastera u višedimenzionalnom prostoru promenljivih. Potom, pošto su klasteri podataka formirani grupisanjem uzoraka, njihova gustina, veličina i oblik mogu značajno varirati. Drugo, posledični deo TSK fazi pravila je opisan linearnom jednačinom koja omogućava razumevanje osnovnih veza između promenljivih. Uprkos tome, jednačina modela sa više promenljivih ima oblik hiper-ravni koju je teško zamisliti ili mentalno vizualizovati. Takođe, kada su prisutni realni podaci na ulazu TSK fazi modela, najčešće se istovremeno aktivira više pravila ovog modela i tada je nemoguće pratiti interakciju između hiper-ravni posledičnih delova pravila [12].

Iz prethodno navedenog je jasno da dva najrasprostranjenija pristupa fazi modelovanja, Mamdani i TSK, odlikuje postizanje dobrih performansi u smislu tačnosti i razumljivosti ponašanja. Mamdani fazi model odlikuju lingvistička fazi pravila koja mogu biti sastavljena direktno na osnovu stručnog znanja eksperta ili pomoću raspoloživih podataka obrađenih izabranom metodom mašinskog učenja

[13]. S druge strane, TSK fazi model je razvijen isključivo na osnovu raspoloživih podataka i predstavlja računarski efikasniji i numerički tačniji model [14]. Oba fazi pristupa modelovanja zasnovana na pravilima odlikuje znatno bolja razumljivost ponašanja i potencijal za analizu, nego što je to slučaj kod modela crne kutije, poput veštačkih neuronskih mreža [15]. Sam pojam interpretabilnosti, odnosno razumljivosti ponašanja se može definisati kao sposobnost izražavanja ponašanja realnog sistema ili modela na razumljiv način [16]. Interpretabilnost obezbeđuje razumevanje na koji način se vrši interakcija strukture i parametara modela (broj pravila, vrednosti lingvističkih promenljivih, oblik fazi skupova, itd.) sa obeležjima podataka, koja obeležja su važna i koje relacije postoje među njima. Često se kaže za razumevanje znanja uskladištenog u modelu, tj. strukture i parametara modela, da predstavlja sveobuhvatnu razumljivost konačnog modela [14]. U cilju dobijanja fazi modela kojeg odlikuje i zadovoljavajuća tačnost i razumljivost ponašanja, za polazni model je odabran TSK fazi model. Polazni model je kombinovan na odgovarajući način sa asocijativnim pravilima u cilju poboljšanja manje razumljivog TSK fazi modela u poređenju sa Mamdani fazi tipom modela [14]. Primenom asocijativnih pravila koja su usklađena sa određenim TSK fazi pravilima, moguće je poboljšati razumljivost ponašanja TSK fazi modela.

Na problemu analize saobraćajnih udesa razvijena je metodologija poboljšanja razumljivosti ponašanja fazi modela na osnovu prikupljenih informacija o stvarnim događajima u saobraćaju (karakteristike udesa, puta i učesnika – personalni podaci). Primena razvijene metodologije je pokazala da je moguće otkriti faktore koji su osnovni uzroci nezgoda i koji su karakteristični za urbane saobraćajnice u gradovima poput Novog Sada. Zato se očekuje da se primenom datog postupka i uključenih modela (TSK fazi pravila i asocijativnih pravila) može obezbediti osnova za poboljšavanje sigurnosti u saobraćaju na osnovu izdvojenih faktora rizika i njihovih skrivenih međusobnih relacija. Drugim rečima, identifikovanje faktora rizika, odnosno faktora koji najčešće utiču na pojavu saobraćajnih nezgoda, kao i njihovih međusobnih zavisnosti, omogućava određivanje najefikasnijeg rešenja za izbegavanje nezgoda ili smanjenja njihovih posledica. Dakle, razvijena metodologija omogućava blagovremen izbor mera čijim uvođenjem se postiže krajnji cilj: da se smanji broj nezgoda u urbanim sredinama i umanjuje uticaj dominantnih faktora rizika.

Kako bi se poboljšala razumljivost ponašanja TSK fazi modela, u predloženi radni okvir je uveden algoritam zasnovan na postupku izdvajanja asocijativnih pravila. Iz inicijalno dobijenog skupa asocijativnih pravila je potrebno izdvojiti

asocijativna pravila koja se podudaraju sa određenim TSK fazi pravilom. Kriterijum podudaranja asocijativnih i TSK fazi pravila je preklapanje između promenljivih koje se pojavljuju u njihovim uzročnim i posledičnim delovima. U toku procesa provere podudarnosti promenljivih, promenljive iz asocijativnih pravila su predstavljene kao fazi singltoni (eng. *singletons*), dok su ulazno-izlazne promenljive TSK fazi pravila aproksimirane Gausovim fazi funkcijama pripadnosti koje su centrirane u odgovarajućim koordinatama centara klastera. Postupak poboljšanja razumljivosti ponašanja za TSK fazi model je razvijen i primenjen na problemu analize saobraćajnih nezgoda u kojima su učestvovali pešaci.

U disertaciji su postavljene sledeće hipoteze:

H1 – Predloženi postupak za selekciju obeležja koji je sastavni deo razvijene metodologije je moguće primeniti na problemima kod kojih nema značajne veze (korelacije) između nezavisnih veličina i izlazne zavisne veličine.

H2 - Predloženi postupak za povećavanje razumljivosti modela zasnovanog na pravilima omogućava izdvajanje najznačajnijih pravila, odnosno pravila koja dominantno utiču na ponašanje modela.

H3 - Predloženi postupak za povećavanje razumljivosti modela zasnovanog na pravilima omogućava transformaciju polaznih fazi pravila TSK forme u jasnijoj i potpunijoj formi AKO-ONDA pravila.

H4 - Predloženi postupak za povećavanje razumljivosti modela zasnovanog na pravilima omogućava vizualizaciju pravila fazi modela koja pomaže boljem razumevanju fazi modela.

### **1.3 Mogućnosti primene**

Primena većine modela zasnovanih na mašinskom učenju je u oblasti predviđanja događaja ili klasifikacije objekata. Pri tome se teži da tačnost predikcije i klasifikacije modela bude što veća, i tada nisu uvek neophodna objašnjenja kako i zbog čega je generisani model izračunao određene vrednosti izlaza. Ipak, u nekim slučajevima je veoma važno da relacije između parametara modela i razmatranih ulaznih promenljivih budu razumljive. Jedna od takvih primena je razumevanje relacija između uzroka i posledica saobraćajnih nezgoda. S obzirom na porast broja nezgoda, pogotovo u urbanim sredinama, razumevanje njihovih uzroka je neophodno prilikom usvajanja i realizacije efikasnih mera za redukovanje broja nezgoda i težine njihovih posledica. Kako su pešaci najugroženija kategorija u saobraćaju u urbanoj sredini, u ovom istraživanju će se kao polazni problem

razmatrati saobraćajne nezgode sa učešćem pešaka. Veoma je važno da se razotkriju interesantne skrivene relacije između promenljivih koje neosporno imaju uticaj na saobraćajne nezgode, kao i na njihovo razumevanje. Takođe je važno da se odredi koji faktori vode ka saobraćajnim nezgodama sa posledicama različite težine.

Iako je dokazano da pomenute tehnike mašinskog učenja daju tačne rezultate pri analiziranju saobraćajnih nezgoda, postignuti rezultati nisu uvek jednostavni za interpretaciju. Nameće se potreba za razvojem metodologije koja će poboljšati razumljivost ponašanja postojećih modela zasnovanih na tehnikama mašinskog učenja.

Postupak za povećavanje razumljivosti ponašanja modela zasnovanog na pravilima, razvijen u okviru ove disertacije, ne zavisi od oblasti primene. Razvijenu metodologiju je moguće primeniti u svim situacijama gde je važno otkrivanje neočiglednih i skrivenih međusobnih uticaja parametara modela.

Metodologija razvijena u okviru disertacije je primenjena u analizi uzroka saobraćajnih nezgoda sa učešćem pešaka. Ovaj tip problema je izabran kao reprezentativni uzorak za prezentaciju razvijene metodologije zbog svog značaja za svakodnevnu bezbednost građana, ali i zbog prisutnih složenih i nejasnih relacija između prisutnih faktora. Na osnovu izdvojenih uticajnih faktora koji utiču na pojavu saobraćajnih udesa kao i njihovih relacija, teži se uvođenju mera predostrožnosti radi smanjenja broja potencijalnih nezgoda u budućnosti, kao i njihove težine. Dakle, pored postojećih predikcionih modela namenjenih za predviđanje saobraćajnih nezgoda, neophodno je i razviti alternativni model sa višim stepenom razumljivosti.

Razvijena metodologija će pomoći razumevanju uzroka saobraćajnih udesa i pronalaženja alternativnih načina analize u cilju poboljšanja bezbednosti učesnika u saobraćaju. Naime, u cilju smanjenja pojave saobraćajnih nezgoda i stepena njihove ozbiljnosti, neophodno je uvesti nove metode analize koje će omogućiti poboljšanje razumljivosti ponašanja modela. Fazi logički modeli generisani tehnikama mašinskog učenja se nameću kao adekvatan izbor. Predloženim istraživanjem je planirano da razvijena metodologija koristi novu metodu analize baze podataka saobraćajnih udesa radi otkrivanja neočiglednih i skrivenih međusobnih uticaja parametara razmatranog modela. Identifikacijom faktora koji utiču na pojavu saobraćajnih nezgoda (vremenske prilike, stanje učesnika u saobraćaju, stanje i karakteristike kolovoza, tj. elemenata puta) se procenjuje rizik njihove pojave.

Najuticajniji faktori rizika se povezuju sa saobraćajnim nezgodama i pronalaze se uzročno-posledične veze između događaja i njegovih uzročnih obeležja.

#### **1.4 Kratak sadržaj disertacije**

Doktorska disertacija je organizovana u šest poglavlja koja su ukratko opisana u daljem tekstu. Prvo poglavlje sadrži uvodna razmatranja iz predmetne oblasti koja se odnosi na poboljšavanje razumljivosti ponašanja fazi logičkih modela. Date su osnovne definicije pojma interpretabilnosti, odnosno razumljivosti ponašanja modela. Nakon pregleda prisutnih problema kod fazi modela, dat je kratak opis razvijenog rešenja i postavljene su hipoteze. Na kraju poglavlja je data kratka diskusija mogućnosti primene razvijene metodologije u analizi uzroka saobraćajnih nezgoda u urbanim sredinama sa učešćem pešaka.

U drugom poglavlju je dat pregled postojećih postupaka i sumiran je pregled literature dosadašnjih istraživanja, kako u oblasti prediktivnih ili klasifikacionih modela zasnovanih na mašinskom učenju, tako i u oblasti interpretabilnosti modela. Težište je stavljeno na primenu prediktivnih modela u oblasti bezbednosti učesnika u saobraćaju.

Treće poglavlje sadrži kratak pregled metoda mašinskog učenja koje su korišćene prilikom izrade ove disertacije, odnosno metode koje su sastavni deo razvijene metodologije. Naglasak je na osnovama postupaka kao što su veštačke neuronske mreže, optimizacija jata jedinki (PSO), grupisanje podataka, fazi logički sistemi i asocijativna pravila.

U četvrtom poglavlju je pregledno opisana celokupna razvijena metodologija, od poboljšanja podataka, preko modelovanja, transformacija razmatranih modela, do analize i verifikacije rezultata. Težište poglavlja je na detaljnom opisu predloženog koncepta povećanja razumljivosti modela zasnovanih na pravilima. U njemu su detaljno opisani koraci ovog postupka, od generisanja različitih predikcionih modela do vizualizacije multi-dimenzionalnih podataka.

Peto poglavlje disertacije sadrži eksperimentalne rezultate dobijene u fazi selekcije obeležja u cilju izbora najpogodnijih varijabli koje će služiti kao ulazi u predikcione modele. Tačnije, dati su rezultati analize i selekcije obeležja sa metodama kros-korelacione analize i hibridne metode koja koristi optimizaciju jata jedinki i veštačke neuronske mreže.

U šestom poglavlju su detaljno izloženi rezultati razvijene metodologije. Rezultati obuhvataju razmatranje dve varijante fazi logičkih modela, a potom

njihovu transformaciju uz pomoć generisanih asocijativnih pravila. Metodologija je ilustrovana na realnom problemu analize saobraćajnih nezgoda u kojima su učestvovali pešaci u gradskom saobraćaju Novog Sada.

U poslednjem, sedmom poglavlju su dati zaključci i obrazloženi osnovni doprinosi doktorske disertacije.





## 2. Pregled aktuelnog stanja u oblasti

Kako bi se rešili sve složeniji problemi u praksi, projektuju se sve sofisticiraniji algoritmi mašinskog učenja, kao što su na primer modeli dubokog učenja. Međutim, takvi složeni modeli nisu razumljivi i bilo kakvo interpretiranje samog modela ili njegovih izlaza je najčešće nemoguće. Danas su sve potrebnije tehnike mašinskog učenja koje omogućavaju generisanje čoveku razumljivih modela ili modela koji mogu obrazložiti svoje odluke ili predikcije. Interpretabilno mašinsko učenje [18] rešava važan problem ljudskog nerazumevanja ponašanja složenih modela mašinskog učenja i doprinosi razumevanju procesa kako takvi modeli donose određene odluke. U ovom poglavlju je dat pregled dva istraživačka pravca. Prvi se odnosi na rezultate istraživanja u oblasti generisanja interpretabilnih fazi logičkih modela. Drugi deo pregleda se odnosi na primene najzastupljenijih tehnika mašinskog učenja u predikciji saobraćajnih udesa, uključujući neuronske mreže, fazi logiku, SVM i hibridne tehnike.

### 2.1 Pregled istraživanja u poboljšavanju interpretabilnosti modela

Takagi-Sugeno-Kang fazi klasifikator zasnovan na slojevitoj strukturi (eng. *stacked-structure*) sa visokom razumljivošću ponašanja predložen je u [19]. Predložena struktura sastoji se od više komponenti predstavljenih u obliku TSK fazi klasifikatora nultog reda. Komponente su slojevito organizovane tako da svaka komponenta pored originalnih ulaza ima i dodatna ulazna obeležja koja odgovaraju izlazima prethodne komponente. Svaka komponenta modela slučajnim izborom ulaznih obeležja i slučajnim izborom pet predefinisanih Gausovih funkcija pripadnosti za odabrana obeležja formira uslovni deo svakog fazi pravila i na taj način garantuje povećanje interpretabilnosti kompletnog modela. Predloženi klasifikator se može obučavati na sličan način kao i neuronska mreža obučena dubokim učenjem (eng. *deep learning network*) koristeći princip slojevite generalizacije (eng. *stacked generalization principle*).

Pomenuti koncept interpretabilnog dubokog TSK fazi klasifikatora (eng. *deep TSK fuzzy classifier*) je dalje poboljšan koristeći pravila sa istim skupom vrednosti lingvističkih varijabli za sve ulaze [20]. Visoka razumljivost ponašanja je postignuta na svakom sloju pravila upotrebom istog skupa lingvističkih vrednosti za

sve ulaze, uključujući i izlaze iz prethodnih slojeva ove složene slojevite strukture. Pokazano je da lingvistička pravila koja za svoje ulaze imaju izlaze iz prethodnih slojeva su ekvivalentna fazi pravilima sa nelinearnim posledičnim delom.

Problem razumljivosti ponašanja TSK modela nultog reda koristeći ulančane manipulacije identifikovane strukture modela redukcijom izlaznih singletona (eng. *singleton*) i uklanjanjem suvišnih pravila predstavljen je u [21]. Razumljivost ponašanja TSK modela je poboljšana simultanom optimizacijom kriterijuma koji uzimaju u obzir transparentnost, složenost i pouzdanost modela. Ovaj cilj je postignut kombinovanjem više poznatih algoritama.

Neuro-fazi mrežna arhitektura koja kombinuje Mamdani i TSK fazi pristupe modelovanja korišćenjem pristupa lokalnog učenja radi postizanja poboljšane razumljivosti ponašanja i tačnosti je predložena u [22]. Ovakav model poboljšava svoju strukturu i parametre primenom algoritma opadajućeg gradijenta (eng. *gradient descent-based algorithm*) i obezbeđuje lingvističku razumljivost ponašanja na izlaznom nivou. Predloženi model odlikuje visoka robustnost, jer se zahvaljujući inkrementalnom učenju adaptira na promenljive karakteristike problema tokom vremena.

Autori su u [23] uspeli poboljšati razumljivost ponašanja TSK fazi pravila uvođenjem novog tipa lingvističkog modifikatora funkcija fazi pripadnosti i novog algoritma učenja. Zahvaljujući tome, predloženi postupak je sposoban da generisanim modelom ispuni simultano dva cilja: poboljšanu interpretabilnost lokalnih linearnih modela i bolju podelu prostora ulaznih obeležja. Pored toga, primenom algoritma prilagođene QR dekompozicije omogućeno je rangiranje fazi pravila.

Postupak generisanja skupa visoko interpretabilnih fazi klasifikatora u formi TSK fazi modela nultog reda, nazvan WL-TSK, predložen je u [24]. Svaki od TSK fazi pod-klasifikatora je generisan zahvaljujući „*drop-out*“ principu regularizacije široko zastupljenom u metodama dubokog učenja. Ovi visoko interpretabilni TSK fazi pod-klasifikatori nultog reda su generisani koristeći postupak generalizacije parametara fazi pravila uvedenog u radu [24] pod nazivom „*fuzzy-knowledge-out*“.

Strukturno evoluirajući pristup je predložen u radu [25]. Ovaj evoluirajući method je zasnovan na inkrementalnom učenju particija podataka. Unapređenje interpretabilnosti je postignuto sa sledećim merama: pronalaskom minimalnog broja fazi pravila, kreiranjem pravila sa minimalnim brojem uslova, i korišćenjem

minimalnog broja aktiviranih pravila potrebnih za dobijanje izlaza modela. Predložena metoda uspešno balansira nivo interpretabilnosti i tačnosti modela.

Autori rada [26] su opisali Python biblioteku koja sadrži komplet softverskih alata sa ciljem poboljšanja interpretabilnosti ekspertnih sistema zasnovanih na fazi kognitivnim mapama kroz poboljšanja u vizualizaciji i reprezentaciji grafova koji se mogu nacrtati koristeći varijable rezultujućih modela.

Struktura učenja zasnovana na konvolucionoj neuronskoj mreži (*CNN*) sa dodatim interpretabilnim slojem u formi fazi logičkih pravila je predložena u [27]. Dodati klasifikacioni sloj je realizovan pomoću RBF neuronske mreže koja predstavlja ekvivalent klasi sistema zasnovanih na fazi logici.

Autori rada [28] predlažu TSK fazi sistem sa zajedničkom i specijalnom bazom znanja (*CSK-TSK-FS*). Parametri koji odgovaraju karakteristikama koji figurišu u posledičnim delovima fazi pravila uvek ostaju nepromenjeni i smatraju se za opšte znanje. S druge strane, specijalno znanje odgovara delovima pravila koji se razlikuju. Pokazano je da je *CSK-TSK-FS* sistem matematički ekvivalentan specijalnom obliku Gausovih kombinovanih modela (eng. *Gaussian mixture model, GMM*).

Zhang i dr. [29] su predložili redukciju broja prostornih ravni ulaznog prostora varijabli pomoću TSK fazi sistema nazvanog DVR-TSK-FS. U okviru DVR-TSK-FS razvijena su sledeća dva mehanizma koji se povezuju sa dubokom strukturom: 1) svakom pogledu u prostoru ulaznih varijabli se dodeljuje odgovarajući obučeni težinski faktor; 2) tokom iteracija DRV-TSK-FS sistema na svakoj komponenti, manje važni pogledi se uklanjaju u skladu za zadatim pragovima.

## **2.2 Pregled istraživanja u projektovanju modela za predikciju saobraćajnih nesreća**

U nastavku ovog poglavlja je dat pregled istraživanja u oblasti analize saobraćajnih udesa pomoću različitih algoritama mašinskog učenja. Algoritmi mašinskog učenja su pokazali da se mogu primeniti na složenim i nepreciznim podacima koji su karakteristični za probleme u vezi sa saobraćajem. Pregled literature koji sledi dokazuje da su se pored fazi logičkih modela i tehnika obrade podataka (eng. *mining techniques*) sa asocijativnim pravilima kao pouzdan alat pokazali i sledeći algoritmi: veštačke neuronske mreže, SVM i hibridni modeli.

### **2.2.1 Modeli zasnovani na veštačkim neuronskim mrežama**

Cilj istraživanja u [30] bio je da se predvide saobraćajne nezgode na putevima u ruralnim područjima Kine. Model za predikciju nesreća kao osnovni faktor koristi geometrijske karakteristike puta na osnovu kojih su deonice podeljene na podsekcije. Prikupljeni su podaci o prosečnom godišnjem dnevnom saobraćaju (eng. *average annual daily traffic, AADT*) i podaci o geometrijskim karakteristikama puteva (radijus horizontalne krive, uzdužni nagib, dužina nagiba, itd). Kao predikcioni modeli su korišćene neuronske mreže sa propagacijom greške unazad (eng. *back-propagation*). Rezultati su pokazali da su najbitnije karakteristike uzdužni nagib, poluprečnik horizontalne krivine i tangenta dužina deonice puta. Model je pokazao da je relativna greška modela predikcije nezgoda u opsegu od 5.85% do 7.87%.

Predikcija saobraćajnih nezgoda je izvršena modelom RBF neuronske mreže u [31]. Ekstrapoliran je broj slučajeva sa fatalnim ishodom i vezani troškovi za period od 2000. do 2006. godine. Za izlazne varijable RBF neuronske mreže su uzete personalne povrede i materijalni gubici kao direktne posledice saobraćajnih nesreća.

Ogrenci [32] predlaže radni okvir zasnovan na neuronskoj mreži u svrhu predikcije težine saobraćajnih nezgoda i akcija koje treba preduzeti nakon registrovanja nesreće. Cilj istraživanja je bio predikcija efekata nesreće na osnovu podataka o prethodnim događajima koristeći neuronsku mrežu, tj. višeslojni perceptron (MLP). Simulacijom je dokazano da MLP modeli uspešno procenjuju težinu ishoda saobraćajne nezgode sa oko 80% tačnosti.

U radu [33] je predložena *Wavelet* neuronska mreža za predikciju materijalnih gubitaka usled saobraćajne nezgode. Neuronska mreža zasnovana na *Wavelet* transformaciji je predložena kao alternativa višeslojnom perceptronu za aproksimaciju proizvoljne nelinearne funkcije. Eksperimentalni rezultati su pokazali da *Wavelet* neuronska mreža poseduje jednostavnu strukturu i brzu konvergenciju.

Istraživanjem u [34] je predstavljen predikcioni model saobraćajnih udesa za raskrsnice korišćenjem neuronske mreže sa propagacijom greške unazad. Ulazne varijable modela su: obim saobraćaja u raskrsnici, lokacija raskrsnice, tip raskrsnice, vrsta puteva koji se spajaju raskrsnicom i režim kontrole saobraćaja, dok izlazna varijabla obuhvata pet vrsta saobraćajnih nezgoda: sa jednim vozilom, sudar od pozadi, sudar od napred, sudar sa strane i sa ogrebotinom. Predloženim modelom je postignuta tačnost predikcije od 89%.

Još jedan primer primene neuronske mreže u svrhu predikcije saobraćajnih udesa opisan je u [35] i predstavljen je neuronskom mrežom obučenom sa *Leveberg-Marquardt* algoritmom propagacije greške unazad. Korišćeni su nacionalni statistički podaci Kine o nesrećama u periodu od 1985. do 1997. godine, a greška predikcije je bila između 0.31% i 2.38%.

U istraživanju [36] je generisan model neuronske mreže sa propagacijom greške unazad na osnovu 1593 zapisa o saobraćajnim nezgodama na 62 raskrsnice sa signalizacijom koji su prikupljeni u gradu Taičung, Tajvan tokom 2000. i 2001. godine. Identifikovane su promenljive od značaja za pojavu saobraćajnih udesa s ciljem predikcije broja udesa na raskrnicama. Pored toga, predložena je šema odlučivanja sa generisanim modelom neuronske mreže, uz pretpostavku da je redukovani broj udesa proporcionalan efikasnosti poboljšanja stanja na raskrnicama, i uz uslov da raskrsnica mora zadovoljiti određene kriterijume da bi bila kandidat za poboljšanje.

Cilj istraživanja [37] bio je da se predvide potencijalne pojave saobraćajnih udesa na autoputu u realnom vremenu, i to identifikovanjem kritičnih i normalnih uslova u saobraćaju. Kritični uslovi u saobraćaju podrazumevaju obrasce u saobraćaju koji vode ka nesrećama, dok su normalni uslovi u saobraćaju uslovi koji ne vode ka nesrećama. Korišćen je klasifikacioni model na bazi PNN (eng. *probabilistic neural network*). PNN model je postigao stopu 86.67% ispravno detektovanih uslova, dok je stopa lažno detektovanih bila 16.67%.

Autori su u radu [38] analizirali saobraćajne nezgode na autoputu u Iranu po mesecima na osnovu obima saobraćaja i stanja putne podloge. Razvijena su tri modela za procenu (prosta regresija, ponderisana regresija i neuronske mreže), a korišćena je i PCA analiza. Rezultati su pokazali da je veštačka neuronska mreža najbolji model za predikciju nesreća sa najmanjom srednjom kvadratnom greškom procene.

U radu [39] je primenjen model negativne binomne regresije i model veštačke neuronske mreže u svrhu analiziranja podataka o saobraćajnim nesrećama u periodu od 1997. do 1988. godine za autoput u Tajvanu. Poređenjem performansi predikcije ova dva modela, zaključeno je da je veštačka neuronska mreža bolja alternativa za analizu frekvencije saobraćajnih udesa na autoputu, sa tačnošću predikcije od oko 60%.

## **2.2.2 Modeli zasnovani na SVM**

Korišćen je SVM (eng. *Support Vector Machine*) kao klasifikator uslova u saobraćaju grupisanih u dve klase: uslove koji vode ka saobraćajnom udesu ili ne [40]. Identifikovani su opasni uslovi u saobraćaju na osnovu prezentovane SVM klasifikacije sa tačnošću od 76.7%.

U okviru rada [41] se koriste SVM modeli za predikciju rizika saobraćajnih nezgoda i sklonosti određenom tipu nezgoda na magistralnim putevima za vozila sa pogonom na dva točka. Predložena metodologija je primenjena na originalnim i na transformisanim vremenskim serijama podataka o saobraćaju u realnom vremenu. Pomenute sklonosti riziku i tipu saobraćajne nezgode su definisane preovlađujućim uslovima u saobraćaju. Za predviđanje rizika saobraćajnih nezgoda, originalne vremenske serije podataka su imale bolje performanse nego transformisane vremenske serije podataka. S druge strane, za predviđanje tipa nezgode su se nešto bolje pokazale transformisane vremenske serije. Rezultati su pokazali da se kombinacija SVM modela i vremenskih serija podataka može iskoristiti u svrhu povećanja bezbednosti na putu, pogotovo pomoću podataka u realnom vremenu.

Autori su u [42] koristili SVM algoritam za procenu rizika saobraćajnih udesa u realnom vremenu. Originalni model je dopunjen inkrementalnim algoritmom kako bi se predikcioni model mogao adaptirati na nove uslove za kratko vreme. Razvijen je predikcioni model koji koristi podatke detektora broja vozila i podatke o saobraćajnim udesima. Testirana su tri eksperimentalna scenarija predviđanja udesa i dokazano je poboljšanje tačnosti modela sa inkrementalnim algoritmom.

## **2.2.3 Modeli zasnovani na pravilima i fazi logici**

U radu [43] je predložen model procene rizika dešavanja saobraćajnih udesa zasnovan na pojednostavljenom modelu hijerarhijske fazi logike (*SHFL*). Pored toga, cilj SHFL modela je bio i da se redukuje broj pravila. Predloženi model karakteriše Mamdani metoda fazi zaključivanja i specifičan modul za heurističko alociranje funkcija pripadnosti. Pored smanjenja broja pravila, sa SHFL modelom se težilo i smanjenju broja nivoa modela hijerarhijske fazi logike.

Cilj sistema odlučivanja zasnovan na Mamdani fazi modelovanju predloženog u [44] je predlaganje odgovarajućih mera za smanjenje saobraćajnih nezgoda. Razvijeni model predviđa prosečan broj saobraćajnih nezgoda koji se može desiti tokom godine, a potom analizom osetljivosti predlaže odgovarajuće

protivmere kojima se suzbijaju uzroci saobraćajnih nezgoda. Na osnovu ovog modela su predloženi kratkoročni i dugoročni planovi za smanjenje udesa.

Primena adaptivnog neuro-fazi sistema zaključivanja (eng. *Adaptive Neuro-Fuzzy Inference System, ANFIS*) za modelovanje saobraćajnih nesreća je opisana u [45]. U poređenju sa nekoliko modela (Poissonova, negativna binomna i nelinearna eksponencijalna regresija), ANFIS model je dao najbolje performanse. Predloženi model se pokazao kao robustan pristup pogodan za modelovanje udesa na osnovu podataka koje odlikuju složene i nejasne relacije.

Postupak grupisanja zasnovan na relacijama fazi ekvivalencija opisan u [46] je primenjen na problemu podele puteva na tri grupe, u zavisnosti od nivoa bezbednosti. Svakoj grupi puteva su dodeljene različite mere za redukciju saobraćajnih udesa.

Autori u [47] su razvili prilagodljiv algoritam za detekciju saobraćajnih udesa zasnovan na fazi tehnologijama. Za ulaze fazi algoritma su koristili brzinu i protok saobraćaja, kao i njihovu varijaciju. Vrednosti fazi lingvističkog ulaza za brzinu su {mala; srednja; velika; veoma velika}, a relevantne vrednosti promene brzine su {mala; srednja; velika}. Algoritam procenjuje dva stanja protoka saobraćaja na osnovu varijacija zauzetosti i protoka. Simulacijama je pokazano da je algoritam imao nižu stopu lažnih procena i višu stopu tačnih u poređenju sa ostalim često korišćenim algoritmima.

U [48] je predložen sistem za analizu saobraćajnih nezgoda zasnovan na Mamdani mehanizmu zaključivanja (eng. *Mamdani inference mechanism*). Ovaj sistem omogućuje identifikaciju stepena izloženosti riziku saobraćajnih udesa. Proučavan je mogući uticaj niza lokalnih kriterijuma za svaki kilometar segmenta analizirane mreže puteva ruralnog područja u severozapadnom Alžiru. Implementiran je niz logičkih pravila koristeći trapezoidne fazi funkcije pripadnosti za opisivanje lingvističkih varijabli. Sistem za evaluaciju meri rizik pojave saobraćajnih nezgoda u cilju poboljšanja nivoa bezbednosti na putevima. Rezultati su pokazali da se razvijeni sistem može uspešno primeniti kao koristan alat u bezbednosti saobraćaja u cilju identifikovanja faktora rizika saobraćajnih udesa vezanih za karakteristike puta.

Autori radova [49] i [50] su konstruisali linearni fazi regresivni model za predikciju saobraćajnih nezgoda. Primenili su postupak fazi linearne regresione analize i postavili su fazi linearno-regresioni model između broja saobraćajnih nezgoda i obima saobraćaja sa različitim vrstama vozila za različite deonice

autoputa u kineskom gradu Čengdu za 2010. godinu. Rešenje modela je dobijeno koristeći *Simplex* postupak. Izlaz fazi relacije je predstavljen prosečnim brojem saobraćajnih nezgoda, a nezavisne varijable su prosečni obimi saobraćaja različitih tipova vozila.

Autori Lin i drugi su u [51] opisali postupak pronalaska pojave najučestalijih obrazaca u saobraćaju na autoputu na osnovu grupisanja obrazaca vrednosti karakteristika saobraćaja i njihove analize. Korišćen je obiman skup podataka prikupljen u državi Virdžinija koji uključuju brzinu i gustinu saobraćaja u različitim vremenskim intervalima. Primenjen je fazi *c-means* postupak grupisanja na svim varijablama. Pokazalo se da vremenske karakteristike, kao što su doba dana ili sedmice, imaju značajan uticaj na saobraćaj.

U istraživanju [52] je primenjen postupak unutrašnjeg proizvoda vektora za fazi grupisanje podataka o saobraćajnim nesrećama u Kini tokom 2010. godine, a uspostavljena je i fazi matrica sličnosti. Na osnovu različitih vrednosti koeficijenata sličnosti iz matrice sličnosti, izvršena je klaster analiza. Rezultati istraživanja su pokazali da je najveća verovatnoća saobraćajnih nezgoda kod automobila, verovatnoća nezgoda kod vozila bez motora i za pešake je srednjeg nivoa, dok je verovatnoća nezgoda za motore najniža.

#### **2.2.4 Modeli zasnovani na asocijativnim pravilima**

Tehnika pretraživanja asocijativnih pravila je još jedan način za pronalaženje skrivenog znanja u podacima koji može biti primenjen u analizi saobraćajnih nezgoda. Prvo su asocijativna pravila generisana iz podataka o nezgodama, a potom su najjača pravila izdvojena na osnovu najviših vrednosti atributa asocijativnih pravila poznatih kao indeksi pouzdanosti i podrške (eng. *confidence and support*) [53].

Slično kao kod prethodnog primera, opsežan skup asocijativnih pravila pronađenih na osnovu podataka o nezgodama je korišćen u [54]. Izdvojena su jaka pravila vezana za indekse pouzdanosti, podrške i devijacije pouzdanosti (eng. *lift*). Pri tome je primenjen *k-means* algoritam grupisanja baziran na proračunu *Jaccard*-ove udaljenosti u cilju određivanja grupa učestalijih pravila.

U istraživanju [55] su primenjene različite metode obrade podataka u svrhu analize saobraćajnih nezgoda, uključujući izdvajanje asocijativnih pravila koristeći Apriori algoritam, grupisanje podataka i redukovanje broja obeležja koristeći u tu svrhu PCA i MDS metode redukcije dimenzionalnosti problema. Studija je pokazala



da je na osnovu raspoloživih podataka i primenjenih metoda analize moguće prepoznati rizične deonice puta i grupe učesnika u saobraćaju koji su najodgovorniji za nastanak udesa u određenim situacijama.

Analiza podataka o saobraćajnim nezgodama u cilju identifikovanja važnijih korelacija između različitih skupova atributa, koristeći pri tome grupisanje podataka i tehnike izdvajanja asocijativnih pravila, je data u [56]. Tehnikom grupisanja je formirano više grupa podataka koje odgovaraju različitim tipovima saobraćajnih nesreća. Dalje, tehnika generisanja asocijativnih pravila je primenjena na pojedinačnim grupama u cilju identifikacije interesantnih pravila koja mogu pomoći u razumevanju uslova pod kojima se nesreće dešavaju.

### 2.2.5 Hibridni modeli i komparacije više modela

Rad [57] opisuje primenu tehnika analize pomoću veštačkih neuronskih mreža (ANN), tačnije višeslojnog perceptrona, i stabla odlučivanja (DT) u svrhu predikcije mesta udesa. Izlazna varijabla predikcionog modela je deonica puta gde se dogodio udes, dok su ulazi: tip vozila, doba dana, vremenski uslovi, itd. Stabla odlučivanja su u svim testovima nadmašile ANN modele. Najbolji rezultat DT modelovanja dobijen je pomoću ID3 algoritma treniranja sa 77.70% ispravno klasifikovanih instanci.

Kombinacija PNN i DT obezbeđuje alternativnu metodologiju za predikciju saobraćajnih nezgoda i opisana je u radu [58]. Cilj je bio istraživanje potencijala kombinovanog modela u predikciji težine udesa (lakši, ozbiljniji, fatalni) za podatke iz policijske baze podataka Kipra za 2005. godinu. Metodologija kojom se kombinuju PNN i DT koristi particije podataka dobijene od najmanjeg neophodnog broja najboljih DT čvorova za treniranje PNN, te se postiže efikasnost i visoka tačnost predikcije za čega su bila dovoljna prva dva DT nivoa.

U [59] je sprovedeno istraživanje sa realnim podacima o saobraćaju koje koristi metod najbližeg suseda (eng. *k-nearest neighbor*, *k-NN*) i metod grupisanja *c-means*. S obzirom da je *k-NN* algoritam nadmašio metodu *c-means* grupisanja, eksperimentalno je pokazano da je *k-NN* algoritam moguće koristiti u svrhu predikcije saobraćajnih nesreća.

Čang i Vang su u [60] prikazali stablo za klasifikaciju i regresiju razvijeno radi analize faktora rizika koji mogu uticati na težinu povreda kao posledice saobraćajnih nezgoda. Stablom odlučivanja se predviđa diskretan izlaz sa vrednostima: bez povreda, sa povredama i sa fatalnim ishodom. Model je obučen i

testiran nad skupom podataka sa 26,831 učesnika u nezgodama u Taipei-ju tokom 2001. godine. Rezultati su pokazali da je tip vozila najbitnija varijabla koja se povezuje sa težinom udesa i da su pešaci, motociklisti i biciklisti identifikovani učesnici u saobraćaju sa većim rizikom da mogu biti povređeni nego ostali učesnici.

Klasifikacija saobraćajnih udesa pomoću ANN i fazi verzije algoritma k-najbližih suseda (k-NN) je proučena u [61], opisujući svaki udes sa 21 varijablom grupisanih zavisno da li se odnose na vozilo, vozača ili uslove vožnje. Performanse oba modela su procenjene kao dobre od strane autora, uz davanje male prednosti ANN klasifikatoru.

U svrhu bezbednosti u saobraćaju, rad [62] daje prikaz tri različita modela za procenu verovatnoće težine posledica saobraćajnih udesa, odnosno povreda i mesta nezgode. Korišćeni su modeli zasnovani na RBF neuronskoj mreži, višeslojnom perceptronu (MLP) i logit modelu. Studija je pokazala da je model RBF neuronske mreže najbolji model za analizu težine povreda vozača, dok je logit model neznatno bolje estimirao lokacije saobraćajnih udesa.

Autori u [63] su istražili primenu MLP i neuronske mreže zasnovane na fazi adaptivnoj rezonantnoj teoriji (fazi ARTMAP) da bi analizirali težinu povreda vozača. Rezultati modelovanja su pokazali da je tačnost klasifikacije na test uzorku za MLP 73.5%, a za fazi ARTMAP 70.6%. Varijable kao što su brzina vozila, pol vozača, korišćenje pojasa, vrsta vozila, i vrsta puta su se pokazali kao faktori koji utiču na ozbiljnost povreda.

U radu [64] su sumirane performanse četiri tehnike mašinskog učenja koje su primenjene na problemu modelovanja težine povreda izazvanih saobraćajnim nezgodama. Primenjene su neuronske mreže obučene pomoću hibridnih postupaka obuke, SVM, stabla odlučivanja (DT) i hibridni model sačinjen od DT i neuronske mreže. Koristeći 10247 zapisa o udesima, cilj je bio da se klasifikuje težina povreda u pet kategorija – bez povreda, moguće povrede, lakše, teške i fatalne povrede. Rezultati su pokazali da je pristup sa hibridnim modelom, DT i neuronske mreže, dao najbolji ishod u odnosu na ostale pristupe.

U radu [65] je korišćena tehnika grupisanja za segmentaciju podataka o saobraćajnim udesima u cilju identifikacije homogenih tipova udesa. Razvijeni su modeli o povredama koristeći multi-nominalni logit model za svaki identifikovani klaster. U istraživanju su korišćeni podaci o saobraćajnim udesima u Belgiji koji opisuju sam udes, vozila, učesnike i karakteristike okruženja. Na osnovu datih karakteristika formirano je sedam klastera. Studijom klaster-modela su

identifikovane varijable koje utiču na težinu povreda. Takođe, ove varijable su upotpunile interpretaciju veze između uslovnih varijabli i ishoda povreda.



### 3. Pregled korišćenih algoritama

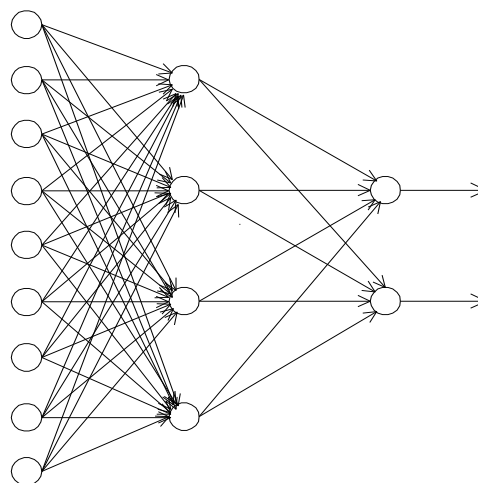
U razvijenom radnom okviru za povećanje razumljivosti modela zasnovanih na pravilima je korišćen veliki broj algoritama mašinskog učenja. Okosnicu radnog okvira čine Takagi-Sugeno fazi model i asocijativna pravila. Tokom procesa generisanja TSK fazi pravila koriste se principi klaster analize realizovani kroz specifičan algoritam samo-organizujuće neuronske mreže. U početnoj fazi radnog okvira se izdvajaju najznačajnija obeležja. U tu svrhu se koristi algoritam optimizacije sa rojem čestica koji je kombinovan sa veštačkom neuronskom mrežom. Veštačka neuronska mreža u formi višeslojnog perceptrona ima ulogu da računa vrednost fitnes funkcije. Pored toga se pri izboru obeležja koristi metoda statističke korelacije. Na kraju radnog okvira se grafički vizualizuje model zasnovan na pravilima. U cilju komparacije razvijenog načina vizualizacije se koristi postupak vizualizacije poznat pod nazivom *Fingrams*. U nastavku poglavlja su objašnjeni osnovni principi navedenih metoda.

#### 3.1 Veštačke neuronske mreže (VNM)

Veštačke neuronske mreže (*eng. artificial neural networks*) predstavljaju oblik modelovanja različitih klasa problema koji može imati niz prednosti u odnosu na konvencionalne numeričke metode. Prednosti veštačkih neuronskih mreža (VNM) se ogledaju u sposobnosti generisanja složenih preslikavanja koje je veoma teško izraziti matematički, velikoj brzini rešavanja problema, robusnosti rada i slično. Veštačke neuronske strukture i algoritmi se razvijaju po analogiji sa poznatim neuro-fiziološkim i neuro-psihološkim procesima učenja [66]. Međutim, današnje strukture i procesi učenja veštačkih neuronskih mreža su još uvek uprošćeni odraz prirodnih neuronskih sistema. Osnovni cilj u razvoju veštačkih neuronskih mreža je težnja istraživača da poznate složene procese koji se odvijaju u mozgu iskoriste za sofisticiranu obradu velikih količina podataka. Glavni zadatak je generisanje algoritama koji imaju sposobnost učenja i prepoznavanja različitih obrazaca.

Višeslojne neuronske mreže su najčešći korišćeni oblik VNM. Sadrže uvek ulazni sloj čvorova - neurona za prijem ulaznih podataka i izlazni sloj neurona, čiji izlazi predstavljaju adekvatne izlaze mreža (Slika 3.1). Mogu imati jedan ili više

procesnih međuslojeva neurona koji se nazivaju skriveni slojevi, jer nemaju neposrednu vezu sa okruženjem VNM. S obzirom da neuroni na istom sloju nisu neposredno povezani, izlazi neurona jednog sloja predstavljaju ulaze u neurone sledećeg sloja. Ova činjenica definiše osobinu višeslojnih neuronskih mreža po kojoj tok signala unutar mreže ide uvek u smeru od ulaznog sloja preko skrivenih slojeva, do izlaznog sloja. U zavisnosti od toga da li su neuroni jednog sloja povezani sa svim neuronima sledećeg sloja ili samo sa nekim od njih, radi se o potpuno ili delimično povezanoj višeslojnoj neuronskoj mreži. Ova klasa neuronskih mreža se često koristi pri rešavanju široke klase problema. Glavne oblasti primene veštačkih neuronskih mreža su u problemima klasifikacije, prepoznavanja oblika i aproksimacije funkcija [67].



Ulazni skriveni i izlazni sloj  
neurona

Slika 3.1 Izgled višeslojne VNM

Veštačke neuronske mreže se generalno klasifikuju u dve velike grupe: VNM sa nadgledanjem (eng. *supervised learning/training*) i VNM bez nadgledanja (eng. *unsupervised learning/training*), u zavisnosti od načina njihovog treniranja [68]. Kada nisu poznati izlazni podaci, za učenje neuronskih mreža se mogu koristiti samo ulazni vektori. Ovakav pristup se zove učenje bez nadzora. Kada su obezbeđeni izlazni i ulazni podaci, takav pristup se zove učenje sa nadzorom. Da bi se izvršila aproksimacija funkcije, klasifikacija uzoraka ili da bi se rešio neki drugi problem pomoću VNM, potrebno je dati problem postaviti u formi skupa primera - uzoraka, koji se naziva obučavajući skup. Specifičnim algoritmima za obuku neuronskih mreža se definišu načini adaptacije težinskih koeficijenata na sinapsama na dovedene različite ulazne uzorke - obučavajuće vektore  $\mathbf{x}$ . Ako je svakom dovedenom ulaznom vektoru  $\mathbf{x}$  pridružen i vektor željenih izlaza  $\mathbf{d}$ , onda metod

podešavanja težina sinapsi radi dobijanja željenog izlaza predstavlja obuku sa nadzorom. Ako je dat samo ulazni vektor i struktura mreže se samo-podešava, onda se radi o obuci bez nadzora. Pored ova dva pristupa obuci veštačkih neuronskih mreža postoje i drugi, manje korišćeni pristupi. Jedan od poznatijih pristupa je obuka sa kritikom (*eng. reinforcement learning*). Kod ovog pristupa se obuka usmerava na bazi indikacije da li neuronska mreža dobro ili loše reaguje na dati ulaz, bez eksplicitnog specificiranja vrednosti željenog izlaza **d**.

Neuronska mreža koja se obučava bez nadzora teži izdvajanju karakteristika ili pravila u posmatranim obrascima. VNM bez nadzora se često koristi za grupisanje podataka, izdvajanje karakteristika ili detekciju sličnosti objekata. Karakterističan predstavnik ovog tipa neuronskih mreža su Kohonenove samo-organizujuće mreže. Najpoznatiji predstavnik VNM kojeg karakteriše obuka sa nadzorom je višeslojni perceptron.

Prethodna podela pristupa obuci VNM je bazirana na prisustvu ili odsustvu određenog tipa „učitelja“ tokom procesa obuke mreže. Osnova za klasifikaciju metoda obuke VNM može biti i na bazi odnosa između uzoraka za obuku i trenutka kada se vrši promena sinapsi. Ako se promena sinapsi vrši nakon obrade svakog uzorka iz skupa za obuku, onda se radi o obuci po uzorku (*eng. pattern-based*). Za razliku od prethodnog pristupa, u *batch* obuci adaptacija težina sinapsi se vrši u jednom mahu na bazi prolaska kroz kompletan ulazno/izlazni skup obučavajućih uzoraka. Algoritamski prolaz kroz sve obučavajuće uzorke se naziva epoha.

### **3.1.1 Neuronske mreže sa nadzorom – Višeslojni perceptroni**

Neuronske mreže višeslojnih perceptrona (*eng. Multilayer Perceptron Neural Networks, MLP*) su često korišćene strukture neuronskih mreža u raznim poljima primene. Neuroni–perceptroni su organizovani u više slojeva i na susednim slojevima su povezani velikim brojem veza-sinapsi (vidi Sliku 3.2) [69]. Ova klasa VNM se sastoji iz tri ili više slojeva neurona. Prvi sloj MLP predstavlja ulazni sloj sastavljen od ‘senzorskih’ jedinica, a poslednji je izlazni sloj mreže. Između ova dva sloja se može nalaziti jedan ili više skrivenih slojeva koji nemaju neposrednu vezu sa okruženjem. Za razliku od ulaznog sloja, neuroni na izlaznim i skrivenim slojevima su procesorskog tipa-perceptroni. Svaki neuron dobija ulazne signale preko zajedničkih sinapsi od čvorova iz prethodnog susednog sloja, a zatim ih prosleđuje prema neuronima na sledećem sloju. Pri tom procesu se ulazni signali u svakom neuronu transformišu putem određene neprekidne i diferencijabilne aktivacione funkcije neurona.

MLP ima dva režima rada: proces obučavanja i eksploatacije. Obučavanje je proces u toku kojeg se vrši korigovanje težinskih koeficijenata na sinapsama između neurona na susednim slojevima u cilju minimizacije greške izlaza mreže. U toku eksploatacije neuronske mreže ulazni signali prolaze u smeru od ulaznog do izlaznog sloja MLP bez korigovanja sinapsi.

Najčešće korišćeni postupak obučavanja neuronskih mreža je postupak prostiranja greške unazad (eng. *backward error propagation, back-propagation, BP*) koji se zasniva na generalizovanom delta pravilu. Postupak pronalaženja vektora gradijenta u strukturi mreže se naziva prostiranje, tj. propagacija greške unazad iz razloga što se gradijentni vektor računa u suprotnom smeru od toka signala u MLP. Ovaj postupak ažurira težinske koeficijente veza (sinapsi) među čvorovima minimizujući zbir kvadratnih greški između željenih i stvarnih izlaza MLP pobuđene datim ulaznim uzorcima. Glavna karakteristika obuke neuronskih mreža prostiranjem greške unazad je dugačko vreme obuke koje se produžava prilikom zahtevnijih složenih mapiranja, ali i kada postoji veći broj skrivenih slojeva.

Razvijene su brojne modifikacije originalne verzije metode prostiranja greške unazad. Proces obučavanja mreže postupkom prostiranja greške unazad ima dve faze-prolaza. U prvoj fazi, težinski faktori na sinapsama se ne menjaju, već se podaci prostiru kroz sve neurone neuronske mreže, od ulaznog do izlaznog sloja. Pri tome se izlaz iz svakog neurona mreže dobija računanjem njegove aktivacione funkcije. Formiranjem razlike između stvarnog krajnjeg izlaza, dobijenog propuštanjem ulaznih podataka kroz neuronsku mrežu, i zadatog ili željenog izlaza, dobija se greška izlaza. Ako se vrednost izlaza svih neurona u poslednjem sloju označi vektorom  $\mathbf{y}$ , a željena vrednost izlaza istih sa  $\mathbf{d}$ , onda se greška izlaza mreže  $E$  može izraziti sa:

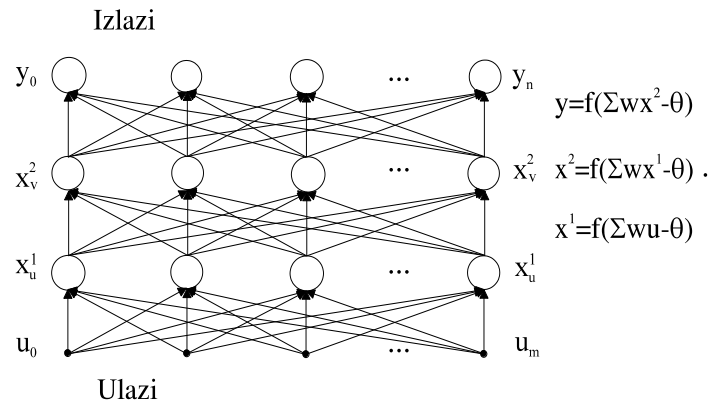
$$E = \frac{1}{2} \sum_i (d_i - y_i)^2. \quad (3.1)$$

U drugoj fazi obuke, algoritam prolazi kroz neuronsku mrežu, sloj po sloj, ali u suprotnom smeru, pri čemu se vrši korekcija težinskih koeficijenata sinapsi. Tokom ovog procesa se metodom negativnog gradijenta vrši minimizacija sume kvadrata greške izlaza  $E$ . Korekcije težinskih faktora sinapsi se tada računaju po obrascu:

$$w'_{ij} = w_{ij} + \Delta w_{ij} \quad (3.2)$$



gde su  $w'_{ij}$  nove, korigovane vrednosti težina sinapsi između  $i$ -tog neurona i proizvoljnih,  $j$ -tih neurona iz prethodnog sloja, a  $\Delta w_{ij}$  je njihova korekcija. Opšti izraz za korekciju težinskih faktora sinapsi je poznat pod nazivom “generalizovano delta pravilo”.



Slika 3.2 Primer VNM sa dva skrivena sloja

Na osnovu prethodno navedenog se može reći da višeslojni perceptron odlikuju tri važne karakteristike:

- Svaki neuron sadrži nelinearnost koja ispunjava uslove neprekidne diferencijabilnosti. U tu svrhu se koristi sigmoidne funkcije, kao što su logistička funkcija (3.3) ili tangens hiperbolična funkcija (3.4);
- MLP neuronska mreža sadrži jedan ili više slojeva skrivenih neurona koji ne pripadaju ulaznim ili izlaznim slojevima mreže. Time se omogućava obuka mreže i njena adaptacija pri rešavanju složenih problema;
- Višeslojni perceptron odlikuje visok nivo povezanosti između neurona u susednim slojevima.

Višeslojna neuronska mreža je adaptivna mreža čiji neuroni imaju istu aktivacionu funkciju. Ova funkcija neurona obično je sastavljena od ponderisane sume i diferencijalne nelinearne aktivacione funkcije, poznate i kao funkcija prenosa neurona. Funkcije aktivacije koje se najčešće koriste su tangens hiperbolične funkcije i logističke funkcije. Obe funkcije, hiperbolična funkcija tangente i logistička funkcija, aproksimiraju funkciju predznaka (eng. *signum*) i odskočnu funkciju, respektivno, i obezbeđuju glatke, nenulte izvode u odnosu na ulazne signale. Krive ovih funkcija prikazuju glatkoću i asimptotska svojstva u obliku slova S, pa se zato još nazivaju i sigmoidne funkcije. Pomenuta sigmoidna funkcija je

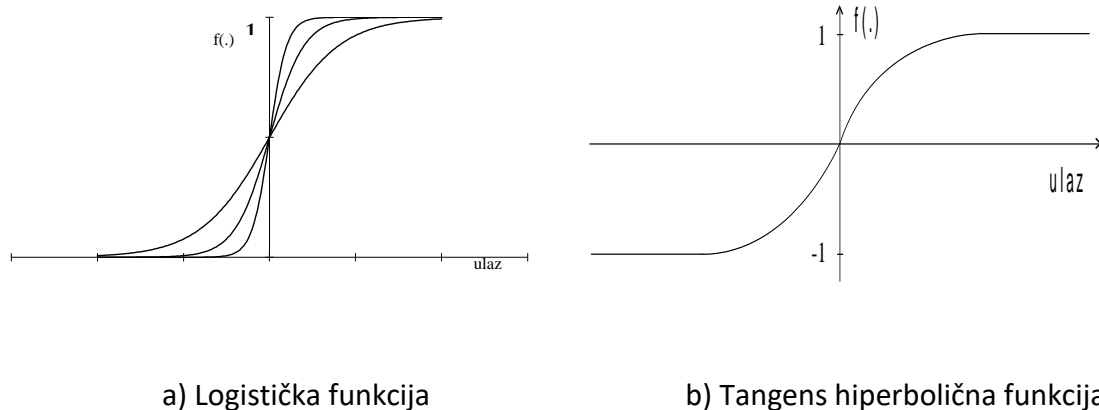
nelinearna, diferencijabilna funkcija i odlikuje se sledećim osobinama: monotono je rastuća, glatka i teži ka asimptotama. Karakterističan oblik logističke funkcije je:

$$f(x) = \frac{1}{1 + e^{-ax}} \quad (3.3)$$

gde je  $a$  parametar nagiba logističke funkcije. Oblik date funkcije je prikazan na Slici 3.3.a) za različite vrednosti parametra nagiba  $a$ , za čije velike vrednosti logistička kriva teži formi jedinične odskočne funkcije. Promenom vrednosti aktivacije neurona u dozvoljenim granicama, izlaz ove funkcije se kreće u intervalu  $[0, 1]$ , te se ova funkcija zove i unipolarna sigmoidna funkcija. U slučaju da je potrebno da izlaz aktivacione funkcije bude u granicama  $[-1, 1]$ , može se koristiti tangens hiperbolična funkcija sledećeg oblika:

$$f(x) = \tanh\left(\frac{x}{2}\right) = \frac{1 - e^{-x}}{1 + e^{-x}} \quad (3.4)$$

Formulom (3.4) je opisana tangens hiperbolična funkcija, odnosno jedna forma bipolarnе sigmoidne funkcije, a njena grafička predstava je data na Slici 3.3.b).



Slika 3.3 Sigmoidne aktivacione funkcije

### 3.2 Samo-organizujuće neuronske mreže

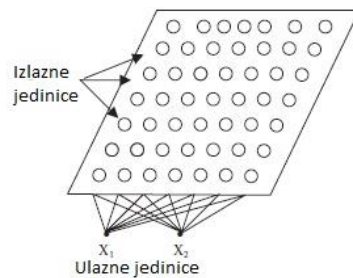
Proces "učenja" kod samo-organizujućih neuronskih mreža se vrši bez nadgledanja [68]. Usled nedostatka informacija vezanih za željene izlaze, neuronske mreže bez nadgledanja ažuriraju težinske koeficijente samo na osnovu ulaznih uzoraka. Samo-organizujuća neuronska mreža se sastoji od ulaznog i izlaznog sloja. Čvorovi ulaznog sloja primaju odgovarajuće elemente vektora karakteristika iz

svakog ulaznog uzorka. Svaki čvor ulaznog sloja je direktno povezan sa neuronima izlaznog sloja. Broj izlaza jednak je broju grupa u koje će se podaci raspoređivati. Pozicija centra grupe je određena vektorom težinskih koeficijenata povezanih sa odgovarajućim izlaznim čvorovima. Upotreba euklidske udaljenosti kao mere razlike ili sličnosti između uzoraka je osnova sheme obučavanja samo-organizujuće neuronske mreže, tj. težine dominantnog čvora se pomeraju prema najbližijem ulaznom vektoru.

### 3.2.1 Kohonen samo-organizujuće neuronske mreže

Kohonen neuronske mreže, poznate još kao Kohonen mape obeležja (eng. *Kohonen feature maps*), pripadaju klasi veštačkih neuronskih mreža sa obukom bez nadzora. Ovaj tip neuronskih mreža se naziva još i topološki organizovane neuronske mape (eng. *topology-preserving maps*), koje imaju snažne asocijacije sa principima funkcionisanja delova biološkog centralnog nervnog sistema [70]. Kohonen neuronska mreža predstavlja skup neurona organizovan u pravilnu strukturu proizvoljne dimenzije. Dvodimenzione Kohonenove mreže (mreže kod kojih se vrši preslikavanje ulaznog vektora visokog reda na neurone raspoređene na dvodimenzionalnu površ) se najčešće primenjuju, ali postoje i Kohonenove mreže sa linijskim organizovanim neuronima ili sa neuronima organizovanim u višedimenzionalnom prostoru. Neuroni su prostorno organizovani u mape koje čuvaju topologiju ulaznog prostora. Na primer, neuroni u određenom delu somatsko-senzorskog dela korteksa odgovaraju na "dodir" susednih prstiju iste ruke. Za Kohonenove mreže važi isto pravilo kao i kod prirodnog, biološkog korteksa, a to je da intenzitet veze između neurona opada sa porastom rastojanja između njih. Dakle, za Kohonenove neuronske mreže važi pravilo da što su neuroni sličniji, to su bliži jedan drugom. Neuroni mogu biti organizovani u ortogonalnu ili heksagonalnu strukturu.

Kohonen neuronska mreža je pogodna za kvantizaciju ulaznog prostora, redukciju dimenzije ulaznog prostora i očuvanje topologije sadržane u strukturi ulaznog prostora. Mreže ovog tipa nameću ograničenje susedstva nad izlaznim jedinicama, poput topološke karakteristike u ulaznim podacima koja se ogleda u težinama izlaznih jedinica. Primer jednostavne Kohonenove samo-organizujuće mreže sa 2 ulaza i 49 izlaza dat je na Slici 3.4.



Slika 3.4 Primer Kohonenove samo-organizujuće mreže

Proces učenja kod Kohonenovih mapa se ogleda u selektovanju mere sličnosti (razlike), a pobjedničkim neuronom se smatra onaj koji ima najveću (najmanju) aktivaciju. U slučaju Kohonenovih mapa karakteristično je da se ažuriraju ne samo težine pobjedničkih neurona, već i sve težine u susjedstvu pobjedničkih neurona. Region susjedstva u mapi se iterativno polako smanjuje tokom procesa učenja.

### 3.3 Algoritmi grupisanja

Tehnikom analize klastera podataka bez nadgledanja se formiraju grupe, tj. klasteri (eng. *cluster*) podataka koji su međusobno slični. Problem klaster analize se definiše kao grupisanje podataka u klaster na takav način da su elementi grupe međusobno sličniji nego što su to sa elementima ostalih klastera. Dakle, algoritmi grupisanja otkrivaju skrivene pravilnosti u velikim i naizgled neorganizovanim skupovima podataka izdvajajući grupe podataka sa karakterističnim zajedničkim osobina. Za klasičan pristup analize klastera važi da svaki uzorak pripada samo jednoj grupi, tj. klasteru. Fazi klaster analiza je generalizacija takvog pristupa, što znači da uzorak može pripadati svim grupama, ali sa različitim stepenom pripadnosti [71, 72].

Primena tehnika klaster analize zavisi od oblasti primene, pa tako postoji nekoliko algoritama za efikasno grupisanje uzoraka poput K-means algoritma i algoritma hijerarhijskog grupisanja [71, 73]. S obzirom na način funkcionisanja postoji više vrsta algoritama klaster analize: hijerarhijski, particioni i algoritmi orijentisani ka gustini klastera. Posljednja vrsta algoritama formira klaster na osnovu njihove gustine, što im omogućava formiranje klastera raznovrsnih formi, dok se prve dve grupe algoritama mnogo češće sreću u literaturi i primeni (obrada slike, bioinformatika, pretraživanje interneta, itd.).

Hijerarhijskim algoritmima se klasteri formiraju rekurzivno, a pri tome se dati klasteri mogu formirati deljenjem većih klastera na manje (eng. *Divisive hierarchical clustering*) ili spajanjem manjih klastera u veće i ovom tipu hijerarhijskih algoritama grupisanja pripadaju aglomerativni algoritmi (eng. *agglomerative hierarchical clustering*). Hijerarhijski pristup rezultuje generisanjem stabla grupisanja (dendrograma) po *bottom-up* strategiji kod aglomerativnih algoritama ili *top-down* strategiji kod deljenja većih klastera na manje. Najčešće se koriste aglomerativni hijerarhijski algoritmi. Aglomerativni algoritmi grupisanja se obično sastoje iz sledećih koraka: (1) Svaki vektor inicijalno predstavlja jednu grupu (jedan vektor – jedna grupa); (2) Izračunavaju se rastojanja između svih grupa koristeći mere sličnosti kao što su Hamingova, Euklidova ili Minkovski distanca; (3) Spajaju se dve najbliže grupe; (4) Algoritam se vraća na drugi korak, sve dok se ne postigne željeni cilj ili ostane samo jedna grupa. Stablo grupisanja - dendrogram je rezultat hijerarhijskih algoritama grupisanja i može se koristiti u interpretaciji strukture podataka i određivanju broja klastera.

Sa particionim algoritmima klaster analize skupovi uzoraka se direktno dele na grupe, gde svaki uzorak pripada jednoj grupi. Iterativnim postupkom grupisanja uzoraka se minimizuje usvojena funkcija cilja, pri čemu broj grupa može biti prethodno definisan ili može biti sastavni deo ciljne funkcije. Particioni algoritam se generalno sastoji iz sledećih koraka: (1) Određivanja broja grupa; (2) Inicijalizacije centara grupa; (3) Računanja usvojene mere grupisanja uzoraka; (4) Ažuriranja koordinata centara klastera; (5) Ukoliko je stanje iz prethodne iteracije nepromenjeno, algoritam se završava, dok se u suprotnom vraća na korak 3.

Ako je broj segmenata nepoznat, particioni algoritam može biti ponovljen na skupu sa drugačijim brojem segmenata, obično u opsegu od dva do  $\sqrt{N}$ , gde je  $N$  broj uzoraka obučavajućeg skupa. Particionim metodama klaster analize pripadaju *K-means*, *fuzzy c-means* i *Gustafson-Kessel* algoritam [74]. Generalno su particione metode bolje od hijerarhijskih jer ne zavise od ranije određenih grupa. Posebno su bolji u situacijama kada su podaci nepravilno raspoređeni. Inače, particione metode implicitno pretpostavljaju oblik grupe podataka. Na primer, *K-means* i fazi *c-means* algoritmi podrazumevaju sferni oblik grupe, dok *Gustafson-Kessel* algoritam omogućava formiranje eliptičnog klastera.

### 3.3.1 *K-means* algoritam klaster analize

*K-means* algoritam grupisanja, poznat još i kao *c-means* algoritam, je iterativni postupak [71, 73]. Svaki klaster poseduje centroid, odnosno centar

klastera koji predstavlja srednju vrednost uzoraka koji pripadaju klasteru. Postupak počinje konfigurisanjem centroida na slučajan način i svaki uzorak se dodeljuje najbližem centroidu. Zadatak je grupisanje datih  $N$  uzoraka od kojih svaki sadrži  $n$  obeležja u  $K$  grupa, na takav način da ciljni kriterijum funkcije grupisanja bude minimalan.

*K-means* algoritam počinje sa zadatim brojem klastera ( $K$ ). U svakoj iteraciji se vrši dodela svakog uzorka onom klasteru čiji je centar najbliži. Centri klastera predstavljaju centroide uzoraka koji pripadaju odgovarajućem klasteru, odnosno centri klastera su zapravo karakteristični predstavnici uzoraka. Nakon što su svi ulazni uzorci dodeljeni nekom klasteru, pozicija centroida svakog klastera se ažurira računanjem srednje vrednosti za klaster, a postupak se ponavlja poboljšavanjem pozicije centroida klastera. Algoritam se završava kada pri prolaženju kroz sve uzorke dođe do situacije da se nije desio ni jedan prelazak nekog uzorka iz jednog u drugi klaster. Alternativno, kriterijum završetka može biti situacija "smirenja" vrednosti kriterijumske funkcije u poslednjih nekoliko iteracija algoritma.

Fazi *c-means* algoritam (FCM) je proširenje *k-means* algoritma i predstavlja najpopularniji fazi algoritam grupisanja [71]. FCM je iterativan postupak kojim se određuje centar klastera u obliku hiper-sfere. Proračun centara klastera se zasniva na minimizaciji indeksa performansi koji sadrži proračun Euklidovog rastojanja između svakog uzorka i svih centara klastera.

### **3.4 Korelacija**

Korelacija je statistička mera za određivanje stepena sličnosti između dve varijable. Odnos između bilo koje dve varijable može varirati od jake do slabe međusobne povezanosti. Mera sličnosti ili povezanosti dve varijable se može izraziti pomoću koeficijenta korelacije. Korelacioni koeficijent može imati vrednosti u opsegu od  $-1.00$  do  $+1.00$ . Kada je korelacioni koeficijent blizu  $+1.00$  (iznad  $r=+.50$ ), stepen sličnosti između varijabli je visok, odnosno smatra se da postoji jaka međusobna pozitivna povezanost (visoka vrednost prve varijable se povezuje sa visokom vrednošću druge varijable, a niska vrednost prve varijable se povezuje sa niskom vrednošću druge varijable). Kako se korelacioni koeficijent približava  $-1.00$  (ispod  $r=-.50$ ), to označava postojanje jake međusobne negativne povezanosti (visoka vrednost prve varijable se povezuje sa niskom ocenom druge varijable, i obrnuto). Korelacioni koeficijent blizak nuli nagoveštava da vrednost jedne varijable ne govori ništa o vrednosti druge varijable, tj. ne postoji međusobna povezanost posmatranih varijabli.

Unakrsna korelacija (eng. *cross-correlation*) se često koristi prilikom proučavanja međusobne zavisnosti dva signala [75]. Unakrsna korelacija kvantifikuje stepen linearne povezanosti između dva stacionarna slučajna procesa  $X(n)$  i  $Y(n)$ , a definiše se kao:

$$R_{xy}(k) = E[X(n+k) \cdot Y(n)], \quad (3.5)$$

gde je  $E$  očekivana vrednost (matematička očekivanost),  $n$  skup celih brojeva, a  $k$  vremenski pomak. Unakrsna korelacija između dva slučajna procesa - ulaznog  $X(n)$  i izlaznog  $Y(n)$  procesa u linearnom sistemu se definiše kao konvolucija:

$$R_{xy}(k) = h(-k) * R_{xx}(k), \quad (3.6)$$

gde je  $h(n)$  impulsni odziv, a  $R_{xx}(k)$  je autokorelacija (unakrsna korelacija signala samog sa sobom).

Autokorelacija vremenski kontinualnog slučajnog procesa  $X(t)$  je definisan kao:

$$R_{xx}(\tau) = E[X(t+\tau) \cdot X(t)], \quad (3.7)$$

gde su vremenske promenljive  $t$  i  $\tau$  definisane nad podskupom realnih brojeva.

Korelaciona matrica ili matrica korelacija  $n$  slučajnih varijabli  $X_1, \dots, X_n$  je  $n \times n$  matrica sa  $i, j$  elementima predstavljenim kao  $corr(X_i, X_j)$ . Simetrija korelacione matrice proizilazi iz ekvivalencije korelacija između  $X_i$  i  $X_j$ , odnosno između  $X_j$  i  $X_i$  varijabli.

### 3.5 Metoda optimizacije sa rojem čestica

Optimizaciju sa rojem čestica (eng. *Particle swarm optimization*, PSO) su prvi opisali J. Kennedy i R. C. Eberhart 1995. godine [76]. Osnovna ideja postupka je proizašla iz asocijacije sa društvenim ponašanjem jedinki iz jata ptica ili riba. PSO je stohastični algoritam pretraživanja koji se zasniva na populaciji adaptivnih rešenja simuliranjem ponašanja jedinki u jatu. Ključni aspekti [77] zbog kojih se smatra da je PSO postao popularan su:

1. Implementacija osnovnog PSO algoritma je relativno jednostavna, s obzirom da se u njegovoj originalnoj verziji primenjuje samo jedan operator za dobijanje novih rešenja.

2. PSO algoritam se pokazao kao veoma delotvoran u mnogim primenama, stvarajući veoma dobre rezultate uz korišćenje niske računarske složenosti [78].

Definicije termina koji se često koriste prilikom opisa PSO algoritma su:

- Jato – populacija rešenja PSO algoritma.
- Čestica – jedinka jata, gde svaka čestica predstavlja potencijalno rešenje problema koji se rešava, a čija pozicija određuje vrednost rešenja koje data jedinka trenutno predstavlja.
- Lokalna najbolja (eng. *personal best*) pozicija čestice – najbolja pozicija čestice koja je dobijena na osnovu izračunate skalarne vrednosti i koja je analogna optimizacionoj funkciji cilja ili fitness funkciji (eng. *fitness*) usvojenoj u teoriji evolucionih algoritmima.
- Globalna najbolja (eng. *global best*) pozicija čestice – najbolja pozicija, tj. rešenje najuspešnije čestice u čitavom jatu.
- Čestica vođa (eng. *leader*) – koristi se za navođenje drugih čestica ka boljim regionima u prostoru pretraživanja.
- Vektor brzine – promenama ovog vektora se odvija proces optimizacije, tj. određuje se pravac kretanja čestice kako bi se poboljšala tekuća pozicija.
- Koeficijent inercije ( $W$ ) – služi za kontrolu uticaja prethodnih brzina na trenutnu brzinu čestice.
- Faktor učenja – određuje nivo privlačenja čestica na osnovu sopstvenog uspeha ili uspeha susednih čestica. Koriste se dva faktora učenja:  $C_1$  (kognitivni) i  $C_2$  (društveni). Faktor  $C_1$  predstavlja privlačenje između čestica na osnovu sopstvenog uspeha čestice, dok faktor  $C_2$  predstavlja privlačenje između čestica na osnovu uspeha susednih čestica. Faktori  $C_1$  i  $C_2$  se obično definišu kao konstante.

Čestice se kreću kroz hiper-dimenzionalni prostor pretraživanja u potrazi za novim rešenjima. Pozicija čestice označena vektorom  $\mathbf{x}_i$  (u trenutku ili iteraciji  $t$ ) se menja dodavanjem vektora brzine  $\mathbf{v}_i$  na trenutnoj poziciji:

$$\mathbf{x}_i(t) = \mathbf{x}_i(t - 1) + \mathbf{v}_i(t) \quad (3.8)$$

Dakle, formulom (3.8) se računa nova pozicija čestice. Vektorom brzine  $\mathbf{v}_i$  se oslikava način na koji čestica odlučuje gde će se pomeriti u sledećoj iteraciji, tj. vremenskom trenutku, uzimajući u obzir prethodne najbolje pozicije, kao i "iskustvo" najuspešnije čestice iz jata. Vektor brzine  $i$ -te jedinke, tj. rešenja je definisan sledećom formulom:



$$\mathbf{v}_i(t) = W\mathbf{v}_i(t-1) + C_1r_1(\mathbf{x}_{pbest_i} - \mathbf{x}_i(t)) + C_2r_2(\mathbf{x}_{gbest} - \mathbf{x}_i(t)), \quad (3.9)$$

gde su  $r_1$  i  $r_2$  nasumične vrednosti u opsegu  $[0,1]$ ,  $W$  je koeficijent inercije, a  $C_1$  i  $C_2$  predstavljaju konstante kojima se kontroliše uticaj lokalnog najboljeg rešenja-čestice  $\mathbf{x}_{pbest}$  i globalno najbolje čestice  $\mathbf{x}_{gbest}$ , respektivno.

Pri svakoj iteraciji PSO algoritma čestice „uče“ jedna od druge i pronalaze novo rešenje na boljoj lokaciji u prostoru rešenja. Svaka čestica pamti lokalno najbolju poziciju ( $pbest$ ) do datog momenta, dok se najbolje rešenje čitavog jata pamti u globalnoj najboljoj poziciji čestice ( $gbest$ ). Ove dve pozicije imaju uticaja na svaku česticu u PSO skupu rešenja, pored trenutnog pravca kretanja svake od čestica.

Pseudo-kod PSO algoritma dat je u Tabeli 3.1. Kriterijum za zaustavljanje algoritma je određen bar jednim od sledećih uslova: maksimalni broj iteracija, određeni broj iteracija bez postizanja poboljšanja kod najboljeg rešenja, ili dostignuta minimalna greška fitnes funkcije.

Tabela 3.1 Pseudo-kod osnovne verzije PSO algoritma

---

**Ulaz:** Nasumično inicijalizovana pozicija i brzina svih čestica:

$X_i(0), V_i(0)$

**Izlaz:** Pozicija aproksimiranog globalnog minimuma  $X^*$

**while** Uslov-zaustavljanja nije zadovoljen **do**

**for**  $i = 1$  to broj čestica **do**

    Računanje fitnes funkcije  $f$

    Ažuriranje lokalne najbolje i globalne najbolje pozicije svake čestice

    Ažuriranje brzine čestice pomoću jednačine (3.9)

    Ažuriranje pozicije čestice pomoću jednačine (3.8)

**end for**

**end while**

---

### 3.6 Asocijativna pravila

Asocijativna pravila su uzročno-posledični iskazi (eng. *if-then statements*) kojima se opisuju međusobne relacije između podataka u bazama podataka. Služe za pronalaženje relacija i njihove važnosti među atributima koji se često pojavljuju zajedno. Dva najčešća i osnovna pojma koji su važna za kvalitet asocijativnih pravila su značaj među atributima (eng. *support*) iz istog skupa podataka, kao mera značaja asocijativnog pravila, i pouzdanost (eng. *confidence*), kao mera tačnosti pravila. Ove dve mere (izražene u procentima) identifikuju veze i pravila generisana analizom podataka i traženjem frekventnih uzročno-posledičnih šablona u podacima [79]. Uobičajeno je da algoritmi za generisanje asocijativnih pravila istovremeno ispune minimalne vrednosti za obe mere, značaj i pouzdanost, zadate od strane korisnika. Još jedna mera kvaliteta asocijativnih pravila koja dodatno opisuju model je „lift“. Ova mera se često koristi pri analizi podataka kao mera devijacije u odnosu na statističku nezavisnost između atributa [80, 81]. Vrednosti ove mere iznad vrednosti jedan ukazuju da postoji pozitivna asocijacija između atributa, a za vrednost jedan ne postoji zavisnost. Postoje još niz mera za ocenu asocijativnih pravila, kao što su npr. *conviction* ili *leverage* [82].

Kriterijum značaja među atributima predstavlja meru frekvencije pojavljivanja atributa u skupu svih opisa događaja, odnosno uzoraka iz baze podataka, a pouzdanost, kao mera jačine asocijativnog pravila, se računa kao što je prikazano formulom (3.10), gde su  $X$  i  $Y$  skupovi uzoraka ( $X$ -uzrok,  $Y$ -posledica),  $X \cap Y = \emptyset$ , a  $N$  je ukupan broj uzoraka u bazi podataka. Dakle, asocijativno pravilo označava da uzrok  $X$  implicira posledicu  $Y$ .

$$\begin{aligned} \text{Značaj među atributima} &= \frac{\text{frekv}(X,Y)}{N} \\ \text{Pouzdanost} &= \frac{\text{frekv}(X,Y)}{\text{frekv}(X)} \end{aligned} \quad (3.10)$$

Pravilo:  $X \Rightarrow Y$

Dakle, iz formule (3.10) se vidi da mera značaja asocijativnog pravila daje relativnu frekvenciju istovremene pojave  $X$  i  $Y$  vrednosti atributa u datom pravilu, dok pouzdanost definiše koliko često je dato pravilo tačno.

Cilj metode formiranja asocijativnih pravila (eng. *association rule mining*) je da se pronađu asocijativna pravila koja zadovoljavaju minimalni predefinisani značaj među atributima i pouzdanost nad posmatranim skupom podataka [83]. Najpre je potrebno pronaći podskup uzoraka (prvi problem) čiji broj ponavljanja premašuje određeni zadati prag (veliki, učestali podskupovi), a zatim je potrebno generisati asocijativna pravila iz prethodno pronađenih učestalih podskupova (drugi

problem), sa ograničenjem da pouzdanost mora biti minimalna [84]. Asocijativna pravila su najčešće generisana na takav način da se istovremeno zadovolji minimalni traženi značaj među atributima i minimalna tražena pouzdanost.

Biće spomenuti samo osnovni, klasični pristupi za pronalazak asocijativnih pravila (eng. *association rule mining*). S obzirom na razumljivost drugog problema, većina istraživanja se uglavnom fokusira na prvi problem, koji se račva na dva podproblema: generisanje kandidata iz velikih podskupova i generisanje učestalih podskupova. Uzimaju se u obzir oni podskupovi čiji značaj među atributima ne premašuje određeni prag i oni podskupovi za koje se očekuje da budu frekventni. Takvi podskupovi se nazivaju „kandidatima“.

Asocijativna pravila i postupci njihovog generisanja se mogu klasifikovati na više načina. U odnosu na dimenzionalnost i slojevitost podataka se generisanje asocijativnih pravila deli na jednoslojna i višeslojna asocijativna pravila, kao i jednodimenzionalna i višedimenzionalna asocijativna pravila. U zavisnosti od načina generisanja asocijativnih pravila postoje modeli za pretragu i generisanje (i) frekventnih podskupova atributa, (ii) sekvencijalnih šablona i (iii) strukturiranih šablona. Na osnovu tipa ograničenja koja se koriste u procesu pretrage i generisanja asocijativnih pravila razlikuju se ograničenja na osnovu tipa znanja, podataka, dimenzionalnosti itd.

Među najpoznatije algoritme pretraživanja i generisanja asocijativnih pravila spadaju: Agrawal- Imielinski- Swami algoritam (AIS) [85], FP-Growth algoritam [86] i Apriori algoritam [83]. AIS [85] algoritam je bio prvi predloženi algoritam za pronalaženje asocijativnih pravila kojim se generišu samo pravila sa samo jednim atributom u posledičnom delu asocijativnih pravila. Nedostatak ovog algoritma je što se generišu uzorci-kandidati za koje se kasnije uspostavi da nisu dovoljno frekventni, a potreban je i velik broj prolaza kroz bazu podataka. *FP-Growth* algoritam je jedan od najefikasnijih algoritama za generisanje asocijativnih pravila predložen od strane Han i dr. [86]. Ovaj algoritam koristi poznatu strategiju '*divide-and-conquer*' prilikom složene pretrage frekventnih podskupova. Postupak se odvija u dve faze. Prvo se svi uzorci iz baze podataka reorganizuju u specifičnu strukturu FP-stabla, a potom se iz njega direktno izdvajaju frekventni podskupovi.

### 3.6.1 Apriori algoritam

Apriori algoritam predložio je Agrawal [83], kao efektivniji način formiranja frekventnih podskupova atributa od kojih se grade asocijativna pravila. S obzirom da Apriori algoritam koristi drugačiji postupak generisanja kandidata i noviju

tehniku redukovanja, znatno je efikasniji algoritam u odnosu na prethodno pomenuti AIS algoritam. Apriori algoritam se koristi za prolanaženje učestalih podskupova uzoraka i „učenje“ asocijativnih pravila pomoću  $k$  podskupova [79]. Razmatra se  $(k+1)$  podskupova i pronalaze najučestaliji podskupovi, odnosno generišu se kandidati. Grupe kandidata su testirane naspram podataka i skupovi se prebrojavaju pomoću pretrage po širini stabla mogućih rešenja (eng. *breadth-first search*) i strukture heš-stabla (eng. *hash-tree*). Identifikuju se pojedinačni frekventni podskupovi iz baze podataka i proširuju se u sve veće podskupove, dokle god se dati podskupovi pojavljuju dovoljno često u bazi podataka. Dakle, Apriori algoritmom se određuju frekventni podskupovi koji se mogu koristiti za određivanje asocijativnih pravila koja ističu generalne trendove u posmatranoj bazi podataka. Nedostaci Apriori algoritma su složenost procesa generisanja uzoraka-kandidata, ali takođe i činjenica da zahteva više prolazaka kroz bazu podataka. Postupak algoritma je dat u Tabeli 3.2, a proces pronalaska kandidata Apriori algoritmom prikazan je na ilustrativnom primeru u Tabeli 3.3.

Tabela 3.2 Pseudo-kod osnovnog Apriori algoritma

---

```
C1k : Podskup kandidata dimenzije  $k$ 
F1k : Frekventni podskup dimenzije  $k$ 
F11 = {frekventni uzorak};
for ( $k=1$ ; F1k != null;  $k++$ )
    C1k+1 = kandidati generisani iz frekventnog podskupa F1k;
    for svaki uzorak  $t$  iz baze podataka  $D$ 
        Inkremenirati brojač kandidata
        C1k+1 sadržani u uzorku  $t$ 
    end for
    F1k+1 = kandidati u C1k+1 sa minimalnim značajem (min_support)
end for
return F1k;
```

---

Kao što se vidi na primeru iz Tabele 3.3, nakon generisanja podskupova uzoraka (eng. *itemset*), baza podataka se skenira i računa se broj uzoraka (frekvencija njihovog pojavljivanja) u svakom podskupu, odnosno značaj među atributima iz istog podskupa podataka koji sadrži podskup traženih atributa. Na osnovu ovih uzoraka i izbacivanjem podskupova čiji je broj uzoraka manji od određenog praga, što zavisi od konteksta (konkretno za ovaj primer je minimalni značaj, tj. *support=4*), generiše se prva grupa selektovanih podskupova (1-

podskupovi) uzoraka (skupovi  $I_1, I_2, I_3, I_5$ -Tabela 3.3(a) i 3.3(b). Samo oni kandidati-podskupovi koji imaju određeni broj uzoraka su generisani i provereni. Kandidati  $k$ -skupovi se generišu nakon svakog  $(k-1)$ -prolaza kroz bazu podataka, pridruživanjem najfrekventnijih  $k-1$ -podskupova. Svi kandidati  $k$ -podskupovi se proveravaju na osnovu njihovih pod- $k-1$ -skupova. Ukoliko se neki od pod- $k-1$ -skupova ne nalazi u listi najfrekventnijih  $k-1$ -podskupova, dati kandidati  $k$ -podskupovi se izbacuju. Ovo se dešava zbog pravila da svaki pod- $k-1$ -podskup najfrekventnijih  $k$ -podskupova mora biti frekventan. Ukoliko se razmatra generisanje podskupova na primeru datom u Tabeli 3.3, uključujući trojke  $(I_1, I_2, I_5), (I_1, I_2, I_3), (I_2, I_3, I_5), (I_1, I_3, I_5)$ , potrebno je proračunati učestanost njihovog pojavljivanja. Potom se ovi podskupovi proveravaju da li sadrže odgovarajuće podskupove. S obzirom da podskup  $(I_3, I_5)$  nije frekventan, parovi podskupova, tj. trojke podskupova  $(I_2, I_3, I_5)$  i  $(I_1, I_3, I_5)$  se eliminišu iz liste kandidata za trojke podskupova (Tabela 3.3(e)). Sa Apriori algoritmom se izbegava nepotrebno prebrojavanje kandidata-skupova za koje se zna da nisu frekventni u dovoljnoj meri.

Tabela 3.3 Ilustrativni primer Apriori postupka

Uzorak	Broj uzoraka	Selektovani 1-podskup uzoraka	Uzorak	Broj uzoraka
$I_1$	7	$I_1$	$I_1, I_2$	5
$I_2$	8	$I_2$	$I_1, I_3$	4
$I_3$	6	$I_3$	$I_1, I_5$	3
$I_4$	2	$I_5$	$I_2, I_3$	4
$I_5$	3	(b)	$I_2, I_5$	3
$I_6$	1		$I_3, I_5$	1

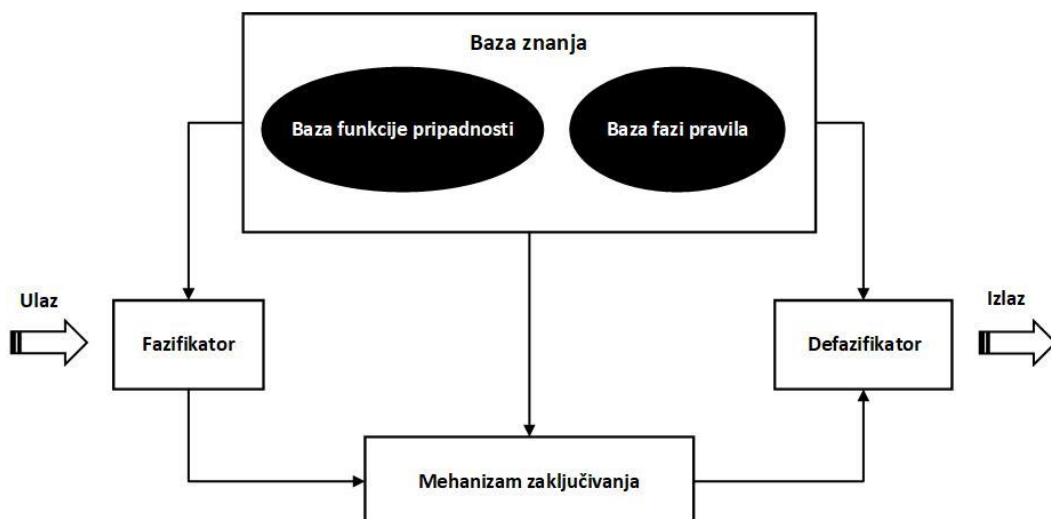
  

Selektovani 2-podskup parova uzoraka	Uzorci	Broj uzoraka
$I_1, I_2$	$I_1, I_2, I_5$	3
$I_1, I_5$	$I_1, I_2, I_3$	2
$I_2, I_5$		
$I_2, I_3$		
$I_1, I_3$		

### 3.7 Fazi logika

Istraživanja u teoriji i praksi, odnosno u primeni teorije fazi skupova (eng. *fuzzy set theory*) profesora Lofti Zadeha, su napredovala od jednostavnijih i specifičnih slučajeva do uobičajenih i kompleksnih sistema [87]. Teorija fazi skupova se u današnje vreme koristi za modelovanje složenih procesa, poput nelinearnih sistema, sistema visokih redova i sistema odlučivanja. Zahvaljujući fazi logičkom modelovanju moguće je rukovanje iskazima opisanim ljudskim jezikom, kao i njihovo konvertovanje u računarske algoritme pogodne za obradu na računaru. Definicija fazi skupa je jednostavno proširenje definicije klasičnog skupa u kojem svaki element može pripadati ili ne nekom skupu. Razlika u odnosu na ovakav konvencionalan skup je što fazi skup opisuje stepen do kog element pripada datom skupu, pa je svakom elementu fazi skupa dodeljen stepen članstva za dati skup sa vrednostima između 0 i 1 [88].

Osnovna struktura bilo kojeg fazi sistema zaključivanja (eng. *fuzzy inference system*) se sastoji od pet komponenti (Slika 3.5): fazifikatora (eng. *fuzzifier*) i defazifikatora (eng. *defuzzifier*) kao sprege ka realnom svetu, gde fazifikator služi za transformisanje realnih varijabli u fazi varijable, a defazifikator vraća fazi varijable u realne varijable. Osnovna baza fazi pravila sadrži odabrana fazi pravila. Funkcije pripadnosti (eng. *membership functions, MF*) baze podataka definišu koje se vrednosti varijabli koriste u fazi pravilima. Centralni deo fazi sistema je mehanizam zaključivanja (eng. *inference engine*), koji obrađuje fazi *if-then* pravila i data ograničenja u cilju generisanja konačnih izlaza ili zaključaka. Ovaj fazi sistem ima svojstva nelinearnog mapiranja prostora ulaznih varijabli u prostor izlazne varijable.



Slika 3.5 Struktura fazi sistema zaključivanja

U cilju boljeg razumevanja daljeg teksta, potrebno je preciznije definisati nekoliko osnovnih pojmova fazi logike kao što su: oblast razmatranja, fazi skup, pripadnost fazi skupu, fazi egzaktni skup (eng. *singleton*), unija fazi skupova, presek fazi skupova, fazi pravila, fazi implikacije i funkcija pripadnosti [89, 90].

- Oblast razmatranja  $\Omega$  predstavlja skup mogućih vrednosti promenljive.
- Fazi skup  $F$  u oblasti razmatranja  $\Omega$  karakteriše funkcija pripadnosti  $\mu_F$ , koja asocira svakoj tački  $\omega$  u  $\Omega$  realni broj u intervalu  $[0,1]$ , odnosno,  $\mu_F : \Omega \rightarrow [0,1]$ :

$$F = \sum_{i=1}^n \mu_F(\omega_i) / \omega_i, \quad (3.11)$$

pri čemu elementi skupa  $\omega$  sa vrednošću  $\mu_F(\omega)$  bliskoj jedinici imaju veći stepen pripadnosti skupu  $F$ .

- Pripadnost fazi skupu je deo oblasti razmatranja  $\Omega$  za koji važi  $\mu_F(\omega) > 0$  i da pripada skupu  $F$ .
- Fazi egzaktni skup je posebna klasa fazi skupa koji ima samo jedan element za koji važi  $\mu_F = 1.0$ .
- Unija fazi skupova  $F$  i  $G$  se označava sa  $F+G$  ili  $F \vee G$  i odgovara "ili" tipu logičke veze. Definisana je sa

$$F + G = \mu_F(\omega) \vee \mu_G(\omega) = \max\{\mu_F(\omega), \mu_G(\omega)\}, \quad \omega \in \Omega. \quad (3.12)$$

- Presek fazi skupova  $F$  i  $G$  se označava sa  $F \cap G$  ( $F \wedge G$ ) i odgovara "i" tipu logičke veze. Presek je određen relacijom:

$$F \cap G = \mu_F(\omega) \wedge \mu_G(\omega) = \min\{\mu_F(\omega), \mu_G(\omega)\}, \quad \omega \in \Omega. \quad (3.13)$$

- Fazi pravilo se može posmatrati kao fazi relacija izražena u formi fazi implikacije. Fazi implikacija se može izraziti preko velikog broja različitih fazi implikacionih funkcija. Za sve je karakteristično da njihovi uzročni i posledični članovi sadrže fazi promenljive i da u sebi sadrže osobine koje su u značajnoj meri intuitivno određene. Fazi implikacione funkcije mogu biti generisane primenom trougaonih normi ili trougaonih ko-normi. T-norma presek za svako  $x, y \in [0, 1]$  ima oblik:

$$T_n(x, y) = x \wedge y = \min\{x, y\}. \quad (3.14)$$

- Najčešće korišćena fazi implikacija je Mamdani fazi implikacija koja koristi presek (min. T-normu):

$$R_m = X \times Y = \int_{U \times V} \mu_X(x) \wedge \mu_Y(y) / (x, y), \quad (3.15)$$

gde su X i Y fazi skupovi koji se nalaze u oblastima U i V, respektivno sa adekvatnim elementima x i y.

- Funkcije pripadnosti se koriste za definisanje fazi skupova. Mogu biti date u numeričkoj ili funkcionalnoj formi. U numeričkoj formi funkcija pripadnosti fazi skupa se predstavlja u obliku vektora brojeva čija je dimenzija određena brojem segmenata oblasti razmatranja. Funkcija pripadnosti proizvoljnog diskretnog skupa ima formu:

$$\mu(x) = \sum_{i=1}^N a_i / x_i, \quad (3.16)$$

gde je  $\mathbf{a}$  vektor oblika  $[a_1 \ a_2 \ \dots \ a_n]$ , što predstavlja numeričku formu funkcije pripadnosti.

Funkcija pripadnosti može biti data u funkcionalnom obliku. Najčešće imaju oblik Gausove krive (3.17) ili njene uprošćene, linearizovane forme (trougla, trapeza i slično). Gausova funkcija pripadnosti (Gausijan) je određena sa dva parametra  $\{\bar{x}, \sigma\}$ , gde  $\bar{x}$  predstavlja srednju vrednost (centar), a  $\sigma$  predstavlja standardnu devijaciju (širinu "osnovice" krive). Gausova kriva je data sa:

$$\mu_A(x, \bar{x}, \sigma) = e^{-\frac{(x-\bar{x})^2}{\sigma^2}}. \quad (3.17)$$

### 3.8 Takagi-Sugeno-Kang fazi-logički modeli

Takagi-Sugeno-Kang fazi logički modeli su predstavljeni metodom učenja na primerima (eng. *learning by examples*) i modelom zasnovanim na fazi pravilima (eng. *fuzzy rule-based model*). Metodom učenja na primerima se generisanje fazi logičkih modela postiže u toku faze identifikacije strukture i identifikacije parametara modela [91, 92]. Potrebno je da fazi logičke modele odlikuju: tačan i brz postupak izračunavanja zahtevane veličine, kao i generalan i fleksibilan postupak, primenjiv na širok spektar najraznovrsnijih problema.



### 3.8.1 Takagi-Sugeno-Kang fazi-logički model

TSK fazi logički model odlikuje se sposobnošću opisivanja složenih, visoko nelinearnih modela, i to korišćenjem malog broja pravila [93, 13], ali je tu prisutan nedostatak u vidu manje razumljive prezentacije znanja. Osnovna ideja TSK fazi logičkog modela zasniva se na činjenici da proizvoljno složeni sistem predstavlja kombinaciju međusobno povezanih podsistema predstavljenih sa  $K$  regiona u razmatranom prostoru stanja. Ponašanje sistema u pojedinim regionima se može opisati jednostavnijim funkcionalnim zavisnostima, koje ukoliko su linearne i ako se svakom regionu dodeli po jedno pravilo, onda TSK fazi logički model može da se predstavi pomoću  $K$  pravila:

$$R_i : \text{If } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \text{ and } \dots \text{ and } x_n \text{ is } A_{in} \\ \text{then } y_i = a_i x + b_i, \quad i = 1, 2, 3, \dots, K, \quad (3.18)$$

gde je  $R_i$   $i$ -to pravilo,  $x_1, x_2, \dots, x_n$  su ulazne varijable,  $A_{i1}, A_{i2}, \dots, A_{in}$  su fazi skupovi dodeljeni odgovarajućim ulaznim promenljivim, promenljiva  $y_i$  predstavlja vrednost izlaza  $i$ -tog pravila, dok su  $a_i$  i  $b_i$  parametri posledične funkcije. Konačan izlaz TSK fazi-logičkog modela za proizvoljni  $x_k$  ulaz se računa pomoću izraza:

$$\hat{y}_k = \frac{\sum_{i=1}^K [\beta_i(x_k) y_i(x_k)]}{\sum_{i=1}^K \beta_i(x_k)}, \quad k = 1, 2, 3, \dots, N, \quad (3.19)$$

gde  $\beta_i$  predstavlja nivo aktivacije (eng. *firing strength*)  $i$ -tog pravila.

Postupak korekcije posledičnih parametara pravila, koji uzima u obzir nivo aktivacije pravila, odnosno tekućih režima (vidi jednačinu 3.19), kao i rezultate globalnog učenja, može uticati na poboljšanje kvaliteta konačnog izlaza modela. Očigledniji opis sistema preko lokalnih modela [94] se može prikazati sledećom formom izraza (3.19):

$$\hat{y}_k = \sum_{i=1}^K [w_i(x_k) (x_k a_i + b_i)] , \quad k = 1, 2, \dots, N, \quad (3.20)$$

gde  $w_{ik}$  predstavlja normalizovani stepen aktivacije  $i$ -tog pravila za  $k$ -ti uzorak. Jednačina (3.21) implicira jednakost izraza (3.19) i (3.20) za konačni izlaz TSK modela:

$$w_{ik} = \frac{\beta_i(x_k)}{\sum_{j=1}^K \beta_j(x_k)}, \quad k = 1, 2, 3, \dots, N. \quad (3.21)$$

TSK fazi-logički model, kao forma skupa lokalnih modela, omogućava korišćenje linearne metode najmanjih kvadrata [75]. Algoritam najmanjih kvadrata

zahteva model koji je linearan po svojim parametrima. Kombinovanje ovako postavljenog globalnog i lokalnog modela učenja se može realizovati putem ponderisane rekurzivne metode najmanjih kvadrata (eng. *weighted recursive least-squares-WRLS*) i pažljivo odabranog faktora zaboravljanja kojim se jednostavno može uravnotežiti uticaji trenutnog stanja i prethodnih stanja [94].

U nastavku sledi opis brzog i tačnog algoritma za generisanje TSK fazi-logičkog modela, primenjivog na široku klasu problema [93, 13]. Algoritam kombinuje grupisanje ulaznih podataka nekom metodom klaster analize, postupak podešavanja parametara metodom najmanjih kvadrata, kao i WRLS procedurom sprovedenom sa istim ciljem. Obuka modela se sastoji iz tri faze. Prva faza predstavlja strukturnu identifikaciju baze pravila. Ovom fazom se vrši grupisanje ulaznog prostora, korišćenjem klaster analize. U sledećoj fazi se vrši identifikacija parametara posledičnog dela skupa pravila. Zadržavajući uslovni deo pravila fiksnim, parametri  $\mathbf{a}_i^T$  i  $b_i$  određuju se metodom najmanjih kvadrata na globalnom nivou. Sa globalno postavljenom metodom najmanjih kvadrata se postiže najmanja greška estimacije parametara modela kompletnog problema. U poslednjoj fazi se vrši adaptacija posledičnih parametara pravila. Ovom fazom se vrši kombinovanje globalnog u lokalnog učenja. Na bazi svakog, novopristiglog k-tog uzorka, metodom WRLS, vrši se adaptacija posledičnih parametara pravila. Pri tome je potrebno izračunati normalizovane nivoe aktivacije svih pravila za dati uzorak  $x_k$ . Na osnovu nivoa aktivacije pravila se određuje u kojem regionu prostora razmatranja se model nalazi. Tada je na osnovu postojećih vrednosti parametara (dobijenih u drugoj fazi), i svakog novog uzorka, moguće WRLS metodom doći do adaptiranih vrednosti parametara  $\mathbf{a}_i$  i  $b_i$  u posledičnom delu pravila. U nastavku slede opisi obe faze obuke fazi-logičkog modela, kao i opšti opis algoritma.

### **3.8.2 Generisanje TSK fazi modela**

#### ❖ Faza I – Strukturna identifikacija modela

Kako bi se ulazno-izlazni prostor ili same ulazne varijable organizovale u više klastera, koristi se samo-organizujuća neuronska mreža (eng. *self-organizing neural network*), odnosno koristi se re-organizujuća neuronska mreža (RONN) koja predstavlja tip samo-organizujuće neuronske mreže sa jednim slojem neurona [95]. U toku njene obuke se uzorci izbalansirano raspoređuju po klasterima i prebacuju iz jednog u drugi klaster, što doprinosi podizanju tačnosti grupisanja. RONN algoritam zahteva da broj klastera bude prethodno definisan. U cilju pronalaženja balansa

između algoritamske složenosti i tačnosti klasifikacije je uveden pojam mrtvog čvora. Mrtvi čvor je centar klastera – prototip koji nije privukao nijedan uzorak.

Svaki čvor izlaznog sloja RONN predstavlja centar jednog klastera. Potrebno je definisati ukupan broj klastera  $C$  kojima će algoritam dodeliti čvorove sa značenjem centara odgovarajućih klastera. Koordinate čvorova mreže su predstavljene pomoću težinskih koeficijenata sinapsi  $w_i$  koji povezuju ulaz mreže i  $i$ -ti čvor. Pretpostavlja se da je zadat neoznačen skup uzoraka  $\{x_1, x_2, \dots, x_N\}$  gde je  $x_k \in R^n$  i  $k=1, \dots, N$ . Opis algoritma obuke RONN neuronske mreže sadrži sledeće korake:

**I korak:** Inicijalizacija jednoslojne NN mreže. Svaki čvor mreže predstavlja centar jednog klastera. Koordinate čvorova NN mreže se inicijalizuju sa slučajnim vrednostima definisanim opsegom ulaznih podataka. Inicijalizacija težina sinapsi svih čvorova se vrši po obrascu:

$$w_{ij} = \left| \min_j x + \left( \max_j x - \min_j x \right) \cdot \rho \right|, \quad i=1, \dots, N; \quad j=1, \dots, n, \quad (3.22)$$

gde je  $\rho$  mali slučajan broj  $\rho \in [0,1]$ .

**II korak:** Koristeći Euklidovo rastojanje pronaći najbliže ulazne uzorke svakom čvoru NN mreže. Za svaki čvor mreže se asocira grupa najbližih ulaznih uzoraka.

**III korak:** Izračunati srednju kvadratnu grešku (SKG) odstupanja uzoraka svakog klastera od njihovog centra:

$$SKG = \frac{1}{n \cdot N_i} \sum_j^{N_i} \sum_k^n (w_{kj} - x_{kj})^2, \quad i=1, \dots, C. \quad (3.23)$$

**IV korak:** Naći aritmetičku sredinu svake koordinate uzoraka grupisanih u svakom klasteru. Rezultat predstavlja nove koordinate centara klastera, odnosno adaptirane težine sinapsi. Za  $i$ -ti čvor će biti:

$$w_i = \frac{\sum_j^{N_i} x_j}{N_i}, \quad i=1, \dots, C, \quad (3.24)$$

gde je  $N_i$  tekući broj uzoraka u  $i$ -tom klasteru.

**V korak:** Proveriti uslov  $\sum_{i=1}^C |w_i - w_i'| \leq \varepsilon$ , gde je  $\varepsilon$  mali pozitivni broj jednak preporučenom iznosu od 10% tekuće vrednosti SKG, a  $w_i$  i  $w_i'$  su vektori sinapsi i-tog čvora u tekućoj i prethodnoj iteraciji. Ako uslov nije ispunjen vratiti se na drugi korak, u suprotnom nastaviti dalje. U slučaju da je u koraku VI detektovano da nema mrtvih čvorova, onda se  $\varepsilon$  postavlja na nulu.

**VI korak:** Identifikovati klaster bez uzoraka, odnosno, mrtve čvorove. Ako takvi postoje izvršiti postupak nalaženja  $K$  klastera ( $K \ll C$ ), za koje su SKG maksimalne. U tu svrhu a priori odrediti broj  $K$  takav da bude jednak preporučenom broju koji iznosi 30% od broja mrtvih čvorova. Potom se mrtav čvor koji predstavlja centar klastera bez uzoraka pomera u blizinu jednog, na slučaj odabranog čvora između  $K$  čvorova - klastera sa maksimalnim vrednostima SKG. Nova lokacija čvora je data formulom:  $w_i^{new} = w_{max}^K + \delta$ ,  $i = 1, \dots, C \wedge Cluster_i = \{0\}$ , gde su:  $w_{max}^K$  lokacija odabranog čvora između  $K$  čvorova sa najvećom SKG,  $w_i^{new}$  nova lokacija centra klastera, a  $\delta = [\delta_1, \delta_2, \dots, \delta_n]$  su mali slučajni brojevi. Ovaj deo koraka VI se ponavlja sve dok se ne odrede nove koordinate svih mrtvih čvorova, odnosno, klastera bez uzoraka. Ako je postignut maksimalan broj iteracija ili ako je u prethodnoj i tekućoj iteraciji broj mrtvih čvorova jednak nuli, onda se algoritam završava. U suprotnom se vraća na korak II.

Pored matrice centara klastera, kao rezultat procesa klasteringa se dobija fazi matrica pripadnosti  $U_{K \times N}$ . Ova matrica sadrži vrednosti  $\mu_{ik} \in [0, 1]$ , za  $1 \leq i \leq K$  i  $1 \leq k \leq N$ , koje predstavljaju stepen pripadanja  $k$ -tog uzorka  $i$ -tom klasteru. Elementi matrice pripadnosti  $U$  se izračunavaju na osnovu sledećeg izraza:

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^K \left( \frac{d_{ik}^2}{d_{jk}^2} \right)^{\frac{1}{(m-1)}}} \text{ za } d_{ik} > 0, \text{ ili } \mu_{ik}^{(l)} = 0 \text{ za } d_{ik} = 0 \quad (3.25)$$

Pri tome, za kolone matrice  $U$  važi uslov:

$$\sum_{i=1}^K \mu_{ik} = 1, \quad k = 1, 2, 3, \dots, N, \quad (3.26)$$

što odgovara uslovu ortogonalnosti fazi skupa ulaznog vektora.

❖ Faza II - Identifikacija posledičnih parametara modela

Faza identifikacije parametara obuhvata postupak izračunavanja parametara TSK fazi-logičkog modela koji figurišu u posledičnom delu pravila. Ključni problem je dobijanje normalizovanog nivoa aktivacije pravila, odnosno, stepena učesća svakog pravila u konačnom izlazu. Zato je prvo potrebno odrediti jednodimenzione fazi skupove  $A_{ij}$  iz uslovnog dela pravila. Dobijaju se projektovanjem multidimenzionalnog fazi klastera u prostor određene ulazne promenljive  $x_j$ , odnosno,

$$\mu_{A_{ij}}(x_{kj}) = proj_j(\mu_{ik}) \quad (3.27)$$

gde je  $\mu_{ik}$  stepen pripadnosti k-tog uzorka (iz vektora  $x_k$ ) i-tom klasteru, a  $\mu_{A_{ij}}(x_{kj})$  je vrednost stepena pripadnosti j-te ulazne veličine k-tog uzorka (j-ta koordinata vektora  $x_k$ ) fazi skupa  $A_{ij}$ . Pošto sve razmatrane funkcije  $\mu_{A_{ij}}(x_{kj})$  predstavljaju funkcije pripadnosti, za njih mora važiti uslov  $\mu_{A_{ij}}(x_{kj}) : R \rightarrow [0, 1]$ . Vrednosti stepena aktivacije i-tog pravila za svaki k-ti ulazni uzorak se računa kao  $\wedge$  - konjunkcija sa operacijom proizvoda:

$$\beta_{ik} = \beta_i(x_k) = \prod_{j=1}^n \mu_{A_{ij}}(x_{kj}), \quad k = 1, 2, 3, \dots, N \quad (3.28)$$

Konačno se za k-ti uzorak normalizovani stepan aktivacije  $w_{ik}$  pravila  $R_i$  dobija po obrascu (3.22).

Sa vrednostima normalizovanog stepena aktivacije dobijenog na osnovu izraza (3.22), formira se k-ti element na glavnoj dijagonali dijagonalne matrice  $W_i$  ( $i=1, \dots, K$ )  $N \times N$  dimenzije. Sada se može formirati matična kompozicija  $\mathbf{X}'$  dimenzije  $N \times K(n+1)$ :

$$\mathbf{X}' = [(W_1 \mathbf{X}_e), (W_2 \mathbf{X}_e), (W_3 \mathbf{X}_e), \dots, (W_K \mathbf{X}_e)], \quad (3.29)$$

gde je matrica  $\mathbf{X}_e = [\mathbf{x}, 1]$  sa vrstama  $[x_k^T, 1]$ . Konačno se parametri  $a_i$  i  $b_i$  iz izraza (3.18) i (3.20), koji figurišu u posledičnom delu svih TSK fazi logičkih pravila, mogu smestiti u vektor  $\theta'$  dimenzije  $K(n+1)$ . U okviru vektora  $\theta'$  parametri su smešteni u niz na sledeći način:

$$\theta' = [\theta_1^T \quad \theta_2^T \quad \dots \quad \theta_K^T]^T, \quad (3.30)$$

gde su  $\theta_i^T = [a_i^T; b_i]$ , za  $i=1, \dots, K$ . Sada se problem dat sa (3.19) ili (3.20) može predstaviti kao regresioni problem  $\hat{\mathbf{Y}} = \mathbf{X}'\theta' + \varepsilon$ , gde je  $\varepsilon$  greška aproksimacije. Vektor nepoznatih parametara  $\theta'$  se može naći metodom najmanjih kvadrata i

obrascom  $\theta' = [(\mathbf{X}')^T \mathbf{X}']^{-1} (\mathbf{X}')^T \mathbf{Y}$ . U praksi se često javljaju slučajevi pravila sa vrlo slabim nivoom aktivacije. Matrica particija tada ima nula kolone ili linearno zavisne kolone, što za finalnu posledicu ima pojavu singularnih ili blisko singularnih vrednosti matrice  $\mathbf{X}'$ . U takvim slučajevima je bolje koristiti metodu dekompozicije singularnih vrednosti (eng. singular value decomposition - SVD), nego Gauss-Newton ili LU dekompoziciju. SVD je numerički robustna i pouzdana ortogonalna transformaciona metoda (Golub [96]).

### 3.8.3 Opis algoritma za generisanje TSK fazi logičkog modela

Algoritam za generisanje TSK fazi logičkog modela predpostavlja da je raspoloživ inicijalni ulazno-izlaznih skup uzoraka za obuku. Na kvalitet izlaza generisanog fazi logičkog modela, pored numeričkih podataka, može značajno uticati izabrani broj generisanih pravila. Sledi opis algoritma obuke TSK fazi modela dat u Tabeli 3.4.

Tabela 3.4 Opis algoritma obuke TSK fazi modela

---

```
// Faza I: Segmentacija prostora varijabli
read skup parova ulazno-izlaznih uzoraka  $(x_k, y_k)$   $k = 1, \dots, N$ ;
set broj pravila  $K$ , stepen rasplinitosti klastera  $m$ ;
call RONN to calculate: rasplinite matrice pripadnosti i matrice centara
    klastera

// Faza II: Identifikacija posledičnih parametara
calculate: stepen aktivacije svakog pravila za sve uzorke  $\beta_{ik} = (\mu_{ik})^n$  i
    normalizovane stepene aktivacije  $w_{ik}$ ;
compose matricu  $\mathbf{X}'$  using dijagonalne matrice normalizovanih stepeni
    aktivacija svih pravila  $\mathbf{W}_i$  i odgovarajućih ulaznih vrednosti;
calculate nepoznati vektor posledičnih parametara  $\theta'$  iz  $\hat{\mathbf{Y}} = \mathbf{X}'\theta' + \varepsilon$  (SVD
    algoritam);
write parametre TSK fazi modela
```

---

## 3.9 Vizualizacija pomoću Fingrams pristupa

*Fingrams* [97, 98] predstavljaju nov način vizualne prezentacije graf - modela koji omogućava jednostavnije proučavanje relacija i svojstava fazi logičkih

modela, ali i ostalih modela zasnovanih na pravilima. Termin *Fingram* je skraćenica izraza *Fuzzy Inference-gram*, a sa formalnog gledišta je definisan kao  $n$ -torka  $(R, P, I, E, m, NSM, NDM)$ , gde:

$R$  je skup fazi pravila (predstavljaju čvorove grafa), označenih sa  $R_i$ ,  $1 \leq i \leq r$ , gde je  $r$  broj pravila;

$P$  je skup fazi particija ulaznih i izlaznih promenljivih;

$I$  je fazi mehanizam zaključivanja;

$E$  je skup raspoloživih uzoraka razmatranog problema, označenih sa  $E_k$ ,  $1 \leq k \leq d$ , gde je  $d$  broj uzoraka;

$m$  je metrika kojom se kreira ponderisana kvadratna matrica  $M$  ( $r \times r$ ) koja predstavlja interakcije među ispunjenim fazi pravilima (pravila čiji su uslovni i posledični delovi ispunjeni u izvesnoj meri). Elementi date matrice su težine vezane za grane (veze) grafa. Koeficijent  $m_{ij}$  je težinski koeficijent veze između pravila  $R_i$  i  $R_j$  koji je definisan jednačinom:

$$m_{ij} = \begin{cases} \frac{SFR_{ij}}{FR_i \cdot FR_j}, & i \neq j \\ 0, & i = j \end{cases}, \quad (3.31)$$

gde  $SFR_{ij}$  odgovara broju uzoraka (instanci) za koja su istovremeno ispunjena pravila  $R_i$  i  $R_j$ , dok  $FR_i$  i  $FR_j$  broje ukupan broj uzoraka za koje su pravila  $R_i$  i  $R_j$  ispunjena, ne uzimajući u obzir da li su ispunjena istovremeno ili ne;

$NSM$  je metoda skaliranja grafa koji se generiše;

$NDM$  je metoda crtanja grafa koji se generiše.

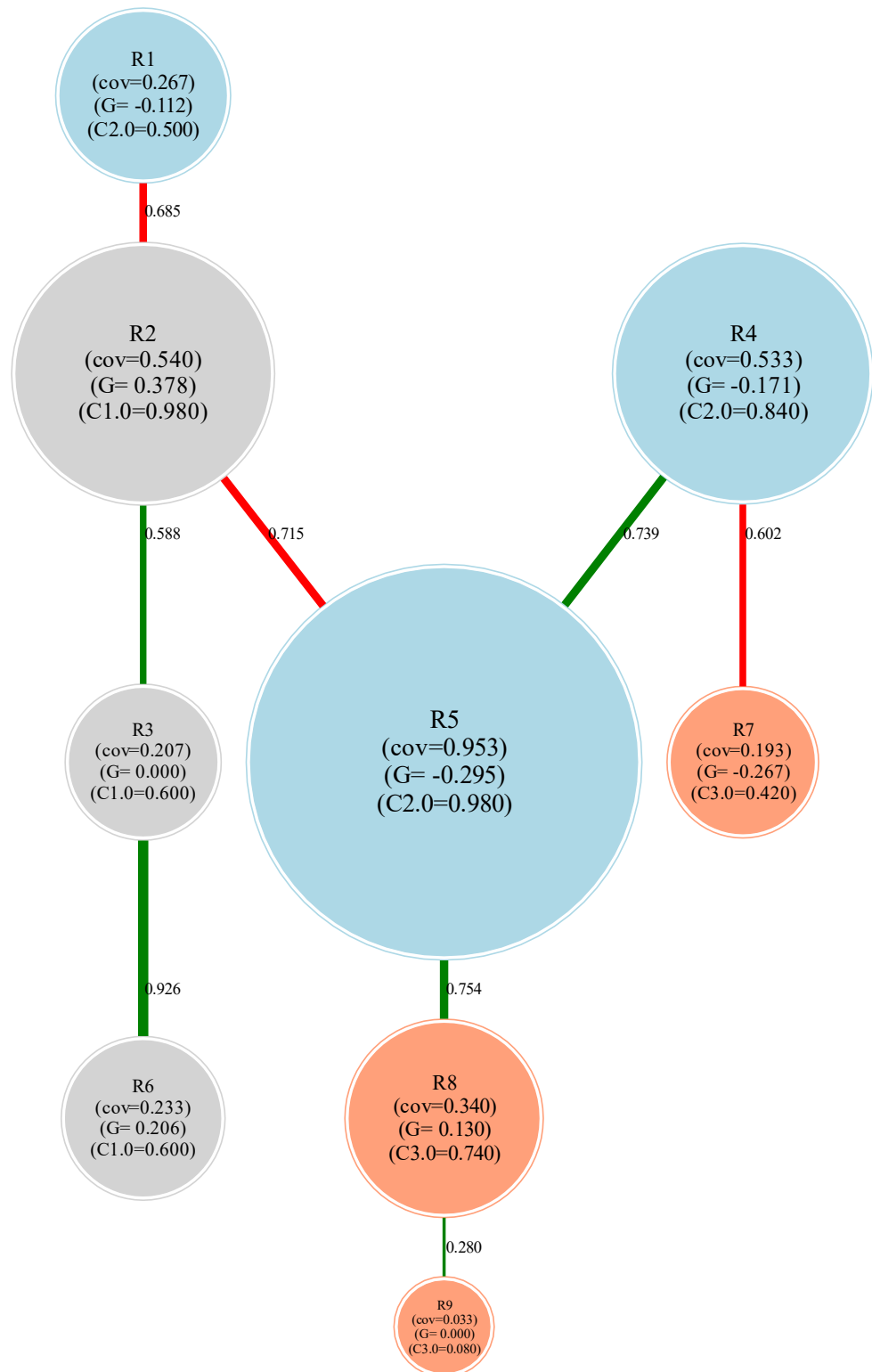
*Fingrams* grafovima se prikazuju fazi modeli u obliku mape logičkog zaključivanja, gde je svako fazi pravilo predstavljeno čvorom datog grafa, dok su relacije između pravila prikazane ponderisanim vezama (granama) čija se vrednost računa pomoću određene metrike. Veličina čvora se uspostavlja na osnovu broja uzoraka koje dato pravilo pokriva. Ivice čvorova označavaju koliko su složeni uslovi datih pravila. Naime, broj ivica koje okružuju čvor predstavljaju broj uslova koji figurišu u uslovnom delu pravila. Na primer, jednolinijska ivica označava samo jedan uslov u pravilu, dvostruka ivica označava dva uslova, itd. Grane, odnosno veze između čvorova predstavljaju informaciju o sličnosti u ispunjavanju uslovno-posledičnih delova pravila. Svaka veza predstavlja relaciju između para fazi pravila. Što je veći nivo preklapanja u ispunjenosti pravila, to je veći težinski koeficijent i debljina linije je veća da bi se veza vizualno jasno predstavila. Pravila koja pripadaju

### *Povećanje razumljivosti modela zasnovanog na pravilima*

istoj klasi prikazana su istom bojom čvorova. Boja veza takođe može biti asocirana za korisne informacije.

Slika 3.6 prikazuje primer grafa sa 9 čvorova koji predstavljaju pravila jednog modela zasnovanog na pravilima gde su jednostavno prikazane najvažnije relacije koje postoje između pravila. Čvorovi u ovom primeru sadrže nekoliko atributa: identifikator pravila ( $R_i$ ), pokrivenost datog pravila ( $cov$ ), relativnu pokrivenost klase izlaza pravila ( $CR_i$ ) i dobrotu pravila ( $G$ ), tj. kako se pravilo ponaša s obzirom na dostupne uzorke problema. Pokrivenost datog pravila  $CR_i$  predstavlja broj uzoraka koji su pokriveni pravilom koje pripada  $i$ -toj klasi podeljen sa ukupnim brojem uzoraka koji odgovaraju istoj klasi. Na osnovu boje čvorova sa date slike se može zaključiti da postoje tri izlazne klase pravila, odnosno tri vrednosti izlazne varijable - *Posledice* (čvorovi su u sivoj, plavoj i crvenoj boji). Broj ivica čvorova pokazuje broj uslova koji figurišu u pravilima. Sva pravila u datom primeru imaju samo jedan uslov što se vidi po broju kružnih ivica oko čvorova. Pored toga, proporcije prečnika čvorova pokazuju relativan odnos pokrivenosti pravila sa uzorcima podataka. Što je broj uzoraka vezanih za pravilo veći, prečnik odgovarajućeg čvora je preporcionalno veći. Na primer, pravilo  $R_5$  ima najviše uzoraka i odgovara drugoj izlaznoj klasi. Konačno, težinski koeficijent koji se nalazi između pravila  $R_3$  i  $R_6$  je najbliži vrednosti 1 (vrednost 0.926) što označava da su ova dva pravila najbližija.





Slika 3.6 Primer Fingrams grafa

Iz prethodno navedenog sledi da veze između čvorova predstavljaju interakciju među pravilima na nivou zajedničkih uzoraka podataka koji ih pokrivaju ukazujući na njihov nivo sličnosti. Što je veći stepen sličnosti između pravila, to je

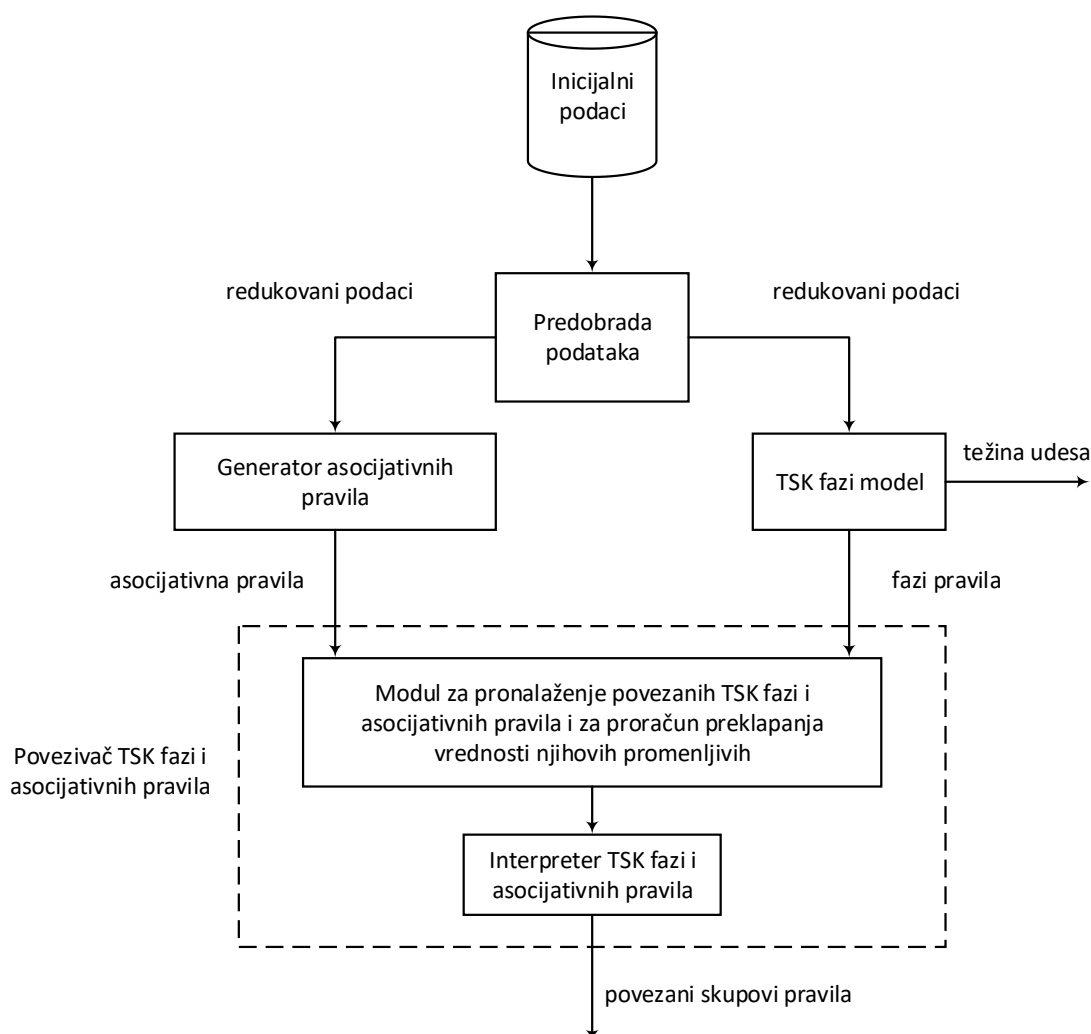
### *Povećanje razumljivosti modela zasnovanog na pravilima*

jači težinski koeficijent veze u grafu. Analiza *Fingrams* grafova omogućava identifikovanje najznačajnijih pravila, detekciju pravila koja pokrivaju iste ili slične uzorke podataka, detekciju potencijalnih nepravilnosti među pravilima i sl. *Fingrams* grafički prikazuju interakciju između fazi pravila na nivou zaključivanja u smislu istovremeno ispunjenih pravila. Njihova grafička predstava nam obezbeđuje da vizualizujemo sažete mrežne modele u dvodimenzionalnom prostoru, što se postiže dodeljivanjem koordinata čvorova sa najviše najjačih veza prema drugim čvorovima u centralnim delovima grafa.

Neki od jednostavnijih, ali korisnih primera primene *Fingrams* analize su: identifikacija pravila koja su pokrivena sa malim brojem uzoraka razmatranog problema, percepcija pravila koja pokrivaju iste uzorke problema, procena razumljivosti modela zasnovanog na pravilima uključujući identifikaciju najbitnijih pravila i detekciju potencijalnih nedoslednosti među pravilima.

## 4. Predloženi radni okvir metodologije

Razvijen je radni okvir za poboljšavanje razumljivosti pravila TSK fazi modela. Predloženi okvir grupiše TSK fazi i asocijativna pravila računajući intervale vrednosti preklapanja varijabli koje se pojavljuju u uzročnim i u posledičnim delovima fazi i asocijativnih pravila. Razvijeni radni okvir obezbeđuje procenu važnosti fazi pravila, relativno u odnosu na druga pravila, ali i u okviru kompletnog TSK fazi modela. Funkcionalnost radnog okvira je ilustrovana na problemu analize saobraćajnih nezgoda sa učešćem pešaka.



Slika 4.1 Shema postupka poboljšanja interpretabilnosti modela

Schema predložene metodologije je data na Slici 4.1. U fazi predobrade se inicijalni podaci čiste i najuticajnije promenljive se selektuju koristeći postupak pretrage prostora rešenja sa optimizacijom roja čestica (eng. *Particle Swarm Optimization, PSO*). Primenjena verzija PSO algoritma koristi višeslojni perceptron (eng. *multilayer perceptron*) za izračunavanje ciljne – fitnes funkcije. Specifične karakteristike ovog PSO algoritma kao i postupak nazvan *Razlike parcijalnih kros-korelacionih matrica* za analizu važnosti varijabli su dati poglavlju 4.1. Korišćeni su podaci sa najznačajnijim obeležjima kako bi se formirala dva nezavisna modela zasnovana na pravilima.

Prvi model koristi TSK fazi pravila, dok je drugi model zasnovan na asocijativnim pravilima. Generator asocijativnih pravila identifikuje najinteresantnije veze između odabranih ulaznih atributa i težine povreda pešaka kao posledica. TSK fazi model generiše skup fazi pravila sa težinama povreda pešaka koristeći iste podatke. S ciljem poboljšanja interpretabilnosti TSK fazi pravila, uveden je povezivač TSK fazi modela i asocijativnih pravila. Ulazi algoritma za povezivanje TSK fazi i asocijativnih pravila su parametri TSK fazi modela i skupa asocijativnih pravila generisanih pomoću Apriori algoritma. Povezivač TSK fazi i asocijativnih pravila se sastoji od dve komponente: Modula za pronalaženje povezanih TSK fazi i asocijativnih pravila, koji potom proračunava preklapanja vrednosti promenljivih za TSK fazi i asocijativna pravila; i Interpretera (eng. *Interpreter*) TSK fazi i asocijativnih pravila. U nastavku poglavlja je dat pregled modula radnog okvira sa posebnim težištem na modul za povezivanje TSK fazi i asocijativnih pravila.

#### **4.1 Predobrada podataka**

Inicijalni skup podataka se sastojao od mnoštva atributa sa potencijalnim uticajem na težinu saobraćajnih nezgoda u kojima učestvuju pešaci. Ovaj skup podataka je najpre očišćen od suvišnih podataka koji su ili nepotpuni ili irelevantni usled nedostajućih ili netačnih vrednosti. Na detaljan proces čišćenja se nadovezuje faza odabira karakteristika ili obeležja. Cilj ove faze je da se odaberu one karakteristike iz prvobitnog skupa karakteristika koje će biti uključene u finalni podskup.

Postoje dva generalna pristupa za selekciju obeležja: „Wrapper“ pristup i pristup filtriranja. „Wrapper“ pristup koristi algoritam mašinskog učenja kojeg iznova poziva tehnika ponovljenog odabiranja (eng. *re-sampling*). Pristupom filtriranja se traže veze među karakteristikama da bi se izdvojile klase, ali i

karakteristike čija je mera povezanosti slabija. U predloženom radnom okviru je primenjena optimizacija roja čestica (PSO) u vidu tehnike pretraživanja prostora karakteristika. Algoritam PSO je zasnovan na socijalnom ponašanju ptica u jatu [76], a koristi roj čestica sačinjen od velikog broja jedinki-čestica u populaciji kandidata podskupova karakteristika. U svakoj PSO iteraciji čestice "uče" jedna od drugih i osvežavaju svoje znanje u odnosu na trenutno i najbolje rešenje, kao i najbolje rešenje celog roja. Svaka čestica u roju je predstavljena svojom pozicijom i brzinom. Pogodnost svake čestice u ovakvoj implementaciji PSO algoritma je procenjena koristeći veštačku neuronsku mrežu (eng. *Artificial Neural Network, ANN*) tipa višeslojnog perceptrona. Algoritam selekcije obeležja zasnovan na optimizaciji roja čestica (PSO) koji je kombinovan sa veštačkom neuronskom mrežom (ANN) će se u daljem tekstu označavati sa PSO+ANN algoritmom. Na ulazu veštačke neuronske mreže je podskup kandidata karakteristika, a trenira se da predvidi ozbiljnost udesa. Izračunata greška predikcije predstavlja pogodnost čestice koja će imati uticaja na trenutnu poziciju čestice (podskup karakteristika), na pravac čestice, a po mogućnosti i na ceo roj. Dakle, korišćeni hibridni postupak selekcije obeležja sa PSO+ANN algoritmom se odlikuje sledećim specifičnim karakteristikama:

- Broj jedinki, ili drugim rečima veličina populacije, određuju koliko će neuronskih mreža biti korišćeno prilikom potrage za najboljom kombinacijom obeležja.
- Svaka neuronska mreža ima drugačiju kombinaciju ulaznih varijabli u jednoj generaciji, ali je njihov broj konstantan.
- Tokom procesa pretrage se težinski koeficijenti sinapsi svih neuronskih mreža iterativno adaptiraju sa ciljem postizanja minimalne greške.
- Postupak pretrage sa hibridnim algoritmom PSO+ANN se završava kada se postigne zadati broj generacija.

Za analizu važnosti pojedinih varijabli i njihove međusobne interakcije korišćen je postupak nazvan *Razlike parcijalnih kros-korelacionih matrica*. Ulaz u algoritam *Razlike parcijalnih kros-korelacionih matrica* je polazni skup podataka i varijabla čija će vrednost biti kriterijum grupisanja. Sledi opis koraka postupka:

Korak 1 - podela skupa uzoraka po vrednostima selektovane varijable - za selektovanu varijablu se uočavaju vrednosti koje može imati i potom se svi uzorci grupišu po kriterijumu da imaju istu vrednost za datu varijablu;

Korak 2 - izračunavanje kros-korelacione matrice za svaku grupu uzoraka;

Korak 3 - za uočene dve grupe uzoraka pravi se razlika kros-korelacionih matrica;

Korak 4 - odabir parova varijabli koji istovremeno imaju bar u jednoj matrici kros-korelacija međusobno visoke vrednosti korelacije i visoku razliku između korelacionih matrica. Granične vrednosti (pragovi) korelacionih matrica se usvajaju empirijski.

## **4.2 Asocijativna pravila i Apriori algoritam**

Cilj pronalaska asocijativnih pravila je da se identifikuju učestali obrasci i veze između promenljivih u skupu podataka posmatranog skupa transakcija. Asocijativno pravilo se može predstaviti implikacijom  $X \Rightarrow Y$ , gde su  $X$  uslovni skupovi obeležja, a  $Y$  je posledično obeležje u transakcijama. Najvažnije mere koje se koriste za ocenu kvaliteta pronađenog asocijativnog pravila su podrška i pouzdanost (eng. *support* ; *confidence*) [84]. Podrška ukazuje na to koliko se često obrasci uslovno-posledičnih obeležja pojavljuju u transakcija. Na pouzdanost se gleda kao na jačinu asocijativnih pravila, koja ukazuje koliko često je pravilo bilo tačno. Cilj pretraživanja asocijativnih pravila je da se pronađu asocijativna pravila koja zadovoljavaju predefinisane minimalnu podršku i pouzdanost [83].

Apriori algoritam [83] je dobro poznat algoritam pretraživanja asocijativnih pravila koji iterativno ponavlja dva osnovna koraka: udruživanje skupova transakcija sa određenim brojem elemenata, tj. obeležja [83] i uklanjanje onih transakcija čija je podrška manja od predefinisane praga [84]. U prvom prolasku kroz skup transakcija, vrednost podrške svake transakcije se računa i veliki skupovi transakcija sa  $l$ -elementa su generisani odbacivanjem transakcija čija je podrška ispod predefinisane minimuma. Skupovi kandidata-transakcija sa istim brojem elemenata su generisani u svakoj iteraciji. Skupovi kandidata sa  $k$ -elementa su generisani nakon  $(k-1)$ -og prolaza pridruživanjem učestalih transakcija sa  $(k-1)$  skupova elemenata. Svaki  $k$ -skup elemenata mora da zadovolji uslov da svaki podskup sa  $(k-1)$  elemenata mora biti učestao. Ovaj proces se nastavlja dok se ne ispostavi da je skup transakcija u tekućoj iteraciji prazan.

## **4.3 TSK fazi model**

TSK fazi model je detaljno objašnjen u prethodnom poglavlju. U daljem tekstu će biti navedene samo osnovne karakteristike TSK fazi modela radi potpunosti opisa metodologije. TSK fazi model opisuje složen sistem u vidu kombinacije međusobno isprepletenih podsistema [99, 93]. Ponašanje svakog

podсистema se može opisati jednostavnijim zavisnostima koje se razlikuju u različitim podsystemima. Algoritam učenja TSK fazi modela sastoji se iz dve faze: identifikacije strukture baze pravila i identifikacije parametara posledičnog dela baze pravila.

Identifikovanje strukture se vrši podelom ulazno-izlazog prostora podataka na klastere. Konkurentni algoritam učenja pod nazivom Re-organizujuća neuralna mreža (*RONN*) koristi se za generisanje klastera podataka [93, 95]. *RONN* predstavlja neuronsku mrežu sa samo jednim slojem, gde svaki čvor predstavlja centar (središte) klastera. U prvom koraku algoritam iterativno podešava koordinate čvorova prema *k-means* algoritmu dok se čvorovi ne stabilizuju na svojim pozicijama [73]. Koordinate čvorova se računaju aritmetički kao uzorci koji pripadaju svakom odgovarajućem klasteru. Srednja kvadratna greška (eng. *Mean Square Error, MSE*) se koristi kao mera devijacije odbiraka (uzoraka) od centra klastera u kojem se nalaze. Centar klastera kojem nisu dodeljeni odbirci (prazan klaster) je pomeren u nasumice odabran proctor, ali u blizini čvora sa najvećom *MSE*, kako bi bio popunjen.

U fazi identifikovanja parametara se računaju prisutni parametri u posledičnom delu TSK fazi pravila [15]. Najvažniji korak je određivanje normalizovane jačine okidanja pravila, odnosno doprinos svakog pravila konačnom izlazu. U tu svrhu, određuje se fazi matrica particija koristeći Euklidsko rastojanje svakog odbirka od centara svih klastera. Svaki element matrice particije predstavlja fazi stepen pripadnosti određenog uzorka datom klasteru. Koristeći matricu particija, moguće je odrediti normalizovanu jačinu aktiviranja svakog pravila na osnovu svakog pojedinačnog uzorka. Kako bi se odredili parametri posledične linearne jednačine, formira se kompozicija matrica. Nju čine normalizovane jačine aktiviranja svih pravila i za sve ulazne uzorke. Kako bi se odredio vektor posledičnih parametara, koristi se metod dekompozicije singularne vrednosti [96].

#### **4.4 Povezivač TSK asocijativnih pravila**

Predloženi postupak koristi prethodno opisane modele TSK fazi i asocijativnih pravila (TSK-AR) koji su generisani nad istim podacima. Ovaj postupak uzima u obzir generalizaciju oblika funkcija pripadnosti za sve ulazno-izlazne promenljive TSK fazi modela sa Gausovom funkcijom. Postupak fazi zaključivanja se vrši između modifikovanih fazi skupova promenljivih TSK modela i promenljivih iz asocijativnih pravila. Formalno, predložena metodologija je definisana kao:

**Definicija:** TSK-AR Povezivač je definisan kao  $n$ -torka (TSK-R, FV, AR, S, FM, V, U, C), gde su:

*TSK-R* – skup TSK fazi pravila  $R_i$ ,  $1 \leq i \leq K$ , definisanih sa (3.18);

*FV* – skup fazi promenljivih  $x_1, x_2, \dots, x_n$  uslovnog dela i fazi izlaza  $y$  posledičnog dela fazi pravila *TSK-R* skupa definisanog funkcijama pripadnosti sa Gausovom formom;

*AR* – skup asocijativnih pravila generisanih nad istim skupom podataka kao *TSK-R* skup;

*S* – skup promenljivih iz *AR* skupa pravila definisanih pomoću singleton funkcija pripadnosti;

*FM* – mehanizam fazi mapiranja korišćen nad fazi promenljivima *FV* skupa i nad promenljivima *S*;

*V* – skup preklapljenih uslovnih delova *TSK-R* i *AR* pravila;

*U* – skup preklapljenih posledičnih delova *TSK-R* i *AR* pravila;

*C* – skup funkcija pokrivenosti *TSK-R* fazi pravila

Ostatak ovog poglavlja objašnjava i detaljnije definiše komponente TSK-AR Povezivača, kao i potrebne korake postupka za kreiranje grafičkog izlaza TSK-AR Povezivača.

*FV* – Skup fazi promenljivih

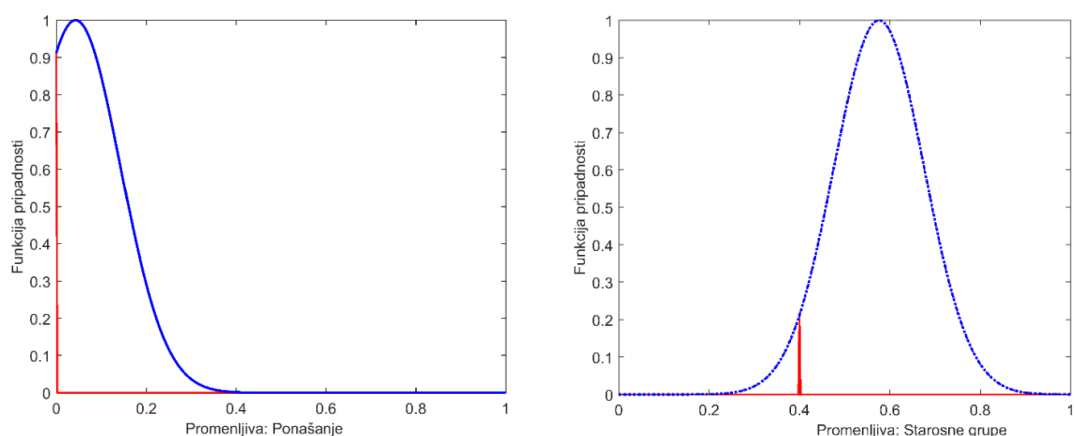
Vrednosti svih fazi promenljivih koje se pojavljuju u TSK-R skupovima su normalizovane u jedinstven prostor oblasti razmatranja u intervalu  $[0,1]$ . Promenljive TSK fazi pravila su promenljive opisane Gausovom funkcijom, tj. funkcijom pripadnosti  $g(x, c, \sigma)$ . Gausova funkcija pripadnosti je definisana pomoću dva parametra: centrom  $c$  i standardnom devijacijom  $\sigma$ . Odgovarajuća koordinata centra klastera za tekuće TSK pravilo je prihvaćena kao centar Gausove funkcije



pripadnosti svake promenljive, dok je za standardnu devijaciju uzeta konstantna vrednost 0.10 iz intervala [0,1] oblasti razmatranja promenljive. Slika 4.2 prikazuje dve Gausove funkcije pripadnosti izdvojene iz uslovnog dela prvog pravila prvog TSK modela (vidi poglavlje 6.2, gde se odgovarajući podaci mogu videti u vrsti Klaster 2 u Tabeli 6.4 i kolonama *Ponašanje* ( $c_1$ ) i *Starosne grupe* ( $c_3$ )). Slika 4.2(a) opisuje funkciju pripadnosti promenljive *Ponašanje* sa centrom u 0.043 (plava kontinualna linija), dok Slika 4.2(b) prikazuje funkciju pripadnosti promenljive *Starosne grupe* sa centrom u 0.576 (plava isprekidana linija). Crvena linija na Slici 4.2 odgovara procesu mapiranja pomenutih funkcija pripadnosti prvog TSK fazi pravila sa odgovarajućim promenljivama iz sedmog asocijativnog pravila (pogledati objašnjenje u opisu komponente: FM - Mehanizam fazi mapiranja).

#### S – Skup promenljivih iz asocijativnih pravila

Vrednosti promenljivih iz skupa asocijativnih pravila ( $AR$ ) se smatraju singleton funkcijama. Neophodno je odrediti poziciju singleton funkcije u okviru unificiranog intervala prostora razmatranja svake promenljive za sva asocijativna pravila. Unificirani interval prostora razmatranja [0,1] svake promenljive je podeljen na  $L-1$  sekcija, gde je  $L$  broj vrednosti lingvističke promenljive. Vrednost promenljive iz  $AR$  skupa asocijativnih pravila određuje raspoloživ skup asocijativnih pravila, što je prikazano u Tabeli 6.2. Za  $l$ -tu lingvističku vrednost singletona  $m$ -te promenljive  $j$ -tog asocijativnog pravila je moguće definisati komponentu označenu kao  $s_{j,m}^l$ . Na primer, druga promenljiva iz Tabele 6.2. (*Pol*), koja može imati vrednosti {Nepoznat, Muški, Ženski}, poseduje jednu od mogućih pozicija singletona {0.0, 0.5, 1.0}. Ukoliko se posmatra promenljiva *Pol* sa značenjem “Ženski” u petom asocijativnom pravilu, odgovarajuća komponenta  $S$  skupa je kodovana sa vrednošću  $s_{5,2}^3 = 1.0$ .



Slika 4.2 Ilustracija preklapanja promenljivih iz skupa fazi promenljivih  $FV$  i skupa promenljivih  $S$

#### FM – Mehanizam fazi mapiranja

Međusobna povezanost TSK fazi pravila i asocijativnih pravila se određuje kao presek singleton funkcija pripadnosti promenljive iz skupa promenljivih  $S$  asocijativnih pravila i Gausove funkcije pripadnosti odgovarajuće fazi promenljive iz skupa fazi promenljivih  $FV$ . Vrednost preseka se dobija izvršavanjem max-min operacije nad ove dve funkcije pripadnosti. Primer delimičnog preseka prvog pravila prvog TSK modela sa sedmim asocijativnim pravilom prikazan je na Slici 4.2. Sedmo asocijativno pravilo je za promenljivu *Ponašanje* singleton funkcije pripadnosti iz skupa  $S$  i Gausove funkcije pripadnosti iz skupa  $FV$  koja ima vrednost 0.912, dok je vrednost max-min operacije za promenljivu *Starosne grupe* 0.213, kao što je prikazano na Slici 4.2(b).

#### V – Skup preklapljenih uslovnih delova pravila

Preklapanje uslovnih delova za  $i$ -to TSK fazi i  $j$ -to asocijativno pravilo  $v_{i,j}$  je definisano kao srednja vrednost maksimalnih vrednosti preseka  $m$ -te promenljive  $j$ -tog asocijativnog pravila i odgovarajuće promenljive  $i$ -tog TSK fazi pravila. Uzimaju se u obzir samo promenljive čiji presek nije prazan skup. Komponenta asocijativnog pravila je predstavljena u obliku singletona  $s_{j,m}^l$  ( $l$ -ta vrednost  $m$ -te promenljive u  $j$ -tom asocijativnom pravilu), dok je odgovarajuća ulazna promenljiva TSK fazi pravila predstavljena Gausovom funkcijom pripadnosti. Dakle, preklapanje  $v_{i,j}$  se računa na sledeći način:

$$v_{i,j} = \sum_m (\maxmin(s_{j,m}^l, g(x_{i,m}; c_{i,m}, 0.1))) / k, \quad (4.1)$$

gde je  $x_{i,m}$   $m$ -ta fazi promenljiva u pravilu  $R_i$  skupa  $TSK-R$ ,  $c_{i,m}$  je  $m$ -ta koordinata centra  $i$ -tog klastera, a  $k$  ( $k \leq n$ ) je broj promenljivih u okviru uslovnih delova pravila sa preklapanjem.

#### U – Skup preklapljenih posledičnih delova pravila

Postupak FM mapiranja se primenjuje na preklapanje posledičnih delova pravila  $TSK-R$  i  $AR$  skupova, na sličan način kao i za uslovne delove pravila, ali uzimajući u obzir da posledični delovi pravila imaju samo jednu promenljivu – *Posledice*. Preklapanje  $u_{i,j}$  posledičnog dela za pravilo  $R_i$  iz  $TSK-R$  skupa i  $j$ -tog asocijativnog pravila iz  $AR$  skupa se određuje kao:

$$u_{i,j} = \maxmin(s_{j,out}^l, g(y_j; c_{j,out}, 0.1)), \quad (4.2)$$

gde je  $s_{j,out}^l$  singleton  $l$ -te vrednosti lingvističke izlazne promenljive u okviru  $j$ -tog asocijativnog pravila,  $y_j$  je izlaz  $j$ -tog fazi pravila iz skupa  $TSK-R$ , a  $c_{j,out}$  je centar

Gausove funkcije pripadnosti koji odgovara vrednosti poslednje koordinate centra  $i$ -tog klastera u ulazno-izlaznom prostoru. Kao što je opisano u poglavlju 3.8.2, klasteri podataka generišu se u ulazno-izlaznom prostoru. Ulazno-izlazni prostor je definisan vektorom promenljivih  $[x^T, y]^T$ , gde  $x$  predstavlja ulazne promenljive, a poslednja koordinata ulazno-izlaznog prostora je izlazna promenljiva  $y$  koja predstavlja promenljivu *Posledice*.

$C$  – skup funkcija pokrivenosti TSK-R fazi pravila

Pokrivenost bilo kojeg  $R_i$  fazi pravila skupa TSK-R asocijativnim pravilima iz AR skupa je definisana pomoću prostora razmatranja AR skupa sa  $r$  asocijativnih pravila ( $r \in AR$ ), odnosno određen je sa brojnošću AR skupa. Funkcija pokrivenosti je generisana na osnovu spajanja, tj. preklapanja više funkcija Gausovog tipa. Centar Gausove funkcije odgovara broju asocijativnog pravila (brojevi 1-7 u poslednjim kolonama Tabele 6.2), dok standardna devijacija odgovara  $v_{i,j}$ , odnosno vrednosti preklapanja uslovnih promenljivih  $j$ -tog asocijativnog i  $i$ -tog TSK fazi pravila definisanog u jednačini (4.1). Gausova funkcija generisana na ovaj način se množi sa  $u_{i,j}$ , t.j. vrednošću preklapanja posledičnih delova  $j$ -tog asocijativnog i  $i$ -tog TSK fazi pravila. Gausove funkcije za sva asocijativna pravila koja imaju preklapanje sa  $R_i$  fazi pravilom se spajaju u funkciju pokrivenosti operatorom unije, odnosno kombinuju se sa operacijom maksimuma. Opis pokrivenosti  $R_i$  fazi pravila sa  $J$  asocijativnih pravilima definiše funkciju pokrivenosti  $C_i$ :

$$C_i = U_i(g(r, j, v_{i,j}) \cdot u_{i,j}), \quad j=1, \dots, J, i=1, \dots, K \quad (4.3)$$

Sve generisane funkcije pokrivenosti  $C_i, i=1, \dots, K$  predstavljaju izlaz TSK-AR algoritma Povezivača. Iscrtavanjem funkcija pokrivenosti omogućuje jednostavnu interpretaciju i bolje tumačenje TSK fazi pravila. Postupak interpretacije je dodatno pojednostavljen ukoliko se razmatraju samo izlazi TSK-AR algoritma Povezivača koji zadovoljavaju unapred zadati prag. Uslov može biti zadat sa maksimalnom vrednošću funkcije pokrivenosti za određeno pravilo koje premašuje predefinisani prag  $Th_1$ , i/ili uzimajući u obzir samo promenljive koje prilikom preklapanja premašuju prag  $Th_2$ .

#### 4.5 Detaljan opis algoritma Povezivača TSK fazi pravila i asocijativnih pravila

Algoritam povezivača TSK fazi i asocijativnih pravila ima za ulaze uzroke i posledice asocijativnih pravila i parametara TSK fazi modela, a prikazan je pseudokodom u Tabeli 4.1. Parametri TSK fazi modela su koordinate centara klastera –

matrice centroida i matrice koeficijenata nagiba. Algoritam poseduje dve glavne petlje, gde spoljašnja petlja iterira kroz listu asocijativnih pravila, dok unutrašnja petlja prolazi kroz TSK fazi pravila. Tokom ovog prolaska kroz TSK fazi pravila prvo se proverava da li je vrednost posledičnog dela  $i$ -tog asocijativnog pravila jednaka poslednjoj koordinati centra klastera koji odgovara  $j$ -tom izlazu TSK fazi pravila. Ukoliko se proverene varijable poklapaju, proverava se da li vrednosti varijabli prethodnog dela  $i$ -tog asocijativnog pravila poklapaju sa odgovarajućom vrednošću koordinate matrice centroida  $j$ -tog TSK fazi pravila. Za pronađeno poklapanje, tj. pogodak, dobija se važnost ulaznih varijabli  $j$ -tog TSK fazi pravila iz  $j$ -tog reda matrice koeficijenata nagiba. Kriterijum pomenutog poklapanja između vrednosti varijabli odgovara opsezima preklapanja kada su vrednosti varijabli bar delimično preklapljene. Na ovaj način se određuju povezana  $i$ -ta asocijativna pravila i  $j$ -ta TSK fazi pravila, koja se smeštaju u listu povezanih pravila. Opsezi varijabli, definisani u opsegu normalizovanih vrednosti od 0 do 1, su podeljeni na ekvidistantne opsege koji odgovaraju definisanim vrednostima obeležja iz Tabele 4.1. Povezana pravila koja su već pronađena se ne ponavljaju.

Tabela 4.1 Algoritam povezivača TSK fazi i asocijativnih pravila

---

```
Get association rules; centroid matrix and slope-coefficient matrix of TSK model
Set  $k$  to 0
for  $i=1$  to number of association rules
  Read association rule( $i$ )
  for  $j=1$  to number of TSK rules
    if consequence-association rule( $i$ ) and centroid( $j$ ,end) are in the overlapping range
      if attributes of causes-of-association rule( $i$ ) and centroid( $j$ ,1 to end-1) are in the overlapped range
        Get slope coefficient( $j$ )
        Create linkedRule from cause's attributes found, consequence and slope coefficients
        if linkedRule is not in the linked rules list
           $k=k+1$ 
          linked rules( $k$ )=linkedRule
          Store linked rules( $k$ )
        end if
```

---

---

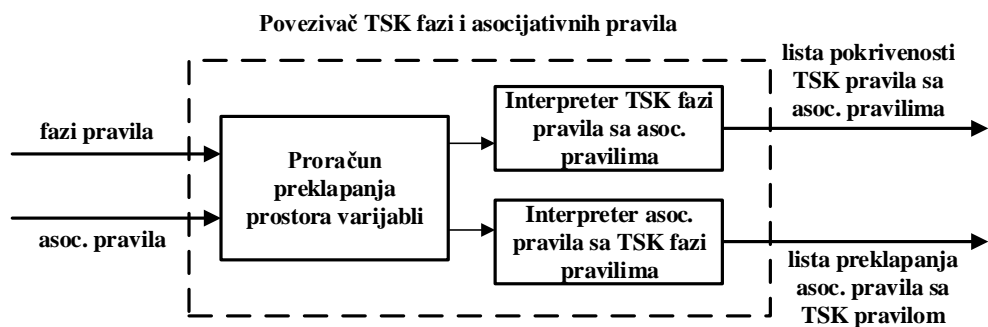
```

end if
end if
end for
end for

```

---

U nastavku teksta je dat opis komponenti algoritma za povezivanje TSK fazi pravila i asocijativnih pravila. Sastoji se od tri komponente: Modula za nalaženje sličnih TSK fazi pravila i asocijativnih pravila, gde se proračunavaju preklapanja prostora varijabli TSK fazi i asocijativnih pravila, Interpretera TSK fazi pravila sa asocijativnim pravilima i Interpretera asocijativnih pravila sa TSK fazi pravilima (vidi Sliku 4.3).



Slika 4.3 Komponente algoritma za povezivanje TSK fazi pravila i asocijativnih pravila

#### 4.5.1 Algoritam interpretiranja TSK fazi modela sa asocijativnim pravilima

Algoritam za interpretiranje TSK modela sa asocijativnim pravilima koristi vektore preklapanja posledičnih delova asocijativnih i TSK fazi pravila i vektor kumulativnog preklapanja svih uslovnih varijabli asocijativnih i TSK fazi pravila (prikazan je pseudo-kodom u Tabeli 4.2). Za svako TSK fazi pravilo se generiše nivo preklapanja njegovog uslovnog dela sa uslovnim delovima svakog asocijativnog pravila. Kao generatorska funkcija nivoa preklapanja se koristi Gausova funkcija raspodele sa centrom koji odgovara rednom broju asocijativnog pravila, a standardna devijacija odgovara srednjoj vrednosti preklapanja uslovnih varijabli asocijativnog i TSK fazi pravila. Tako generisana Gausova funkcija se množi sa vrednošću preklapanja posledičnog dela asocijativnog i TSK fazi pravila. Vrednosti Gausovih funkcija preklapanja za sva asocijativna pravila koja imaju preklapanje sa tekućim TSK fazi pravilom se kombinuju sa funkcijom maksimuma. Rezultat

kombinovanja u celom intervalu preklapanja ne može imati veću vrednost od maksimalne vrednosti jednake jedinici. U nastavku je dat pseudo-kod algoritma za interpretiranje TSK modela sa asocijativnim pravilima. Promenljiva *ConsOverlap* predstavlja preklapanja posledičnih delova svih parova asocijativnih i TSK fazi pravila, dok promenljiva *CumulCausesOverlap* ima značenje kumulativnog preklapanja svih uslovnih varijabli svih parova asocijativnih i TSK fazi pravila.

Tabela 4.2 Algoritam interpretiranja TSK fazi modela sa asocijativnim pravilima

---

<pre><b>Get</b> vector ConsOverlap; vector CumulCausesOverlap <b>Init</b> TSKruleCover <b>for</b> <i>i</i>=1 to number of TSK rules   <b>for</b> <i>j</i>=1 to number of association rules     <b>if</b> CumulCausesOverlap&gt;0       TSKruleCover(<i>i</i>)=<b>max</b>(TSKruleCover(<i>i</i>),         <b>Gauss</b>(center:<i>j</i>,stdDev:<b>mean</b>(CumulCausesOverlap(<i>i</i>,<i>j</i>)))       <span style="float:right">×</span>     ConsOverlap(<i>i</i>,<i>j</i>)     <b>end if</b>   <b>end for</b> <b>end for</b></pre>
--

---

#### 4.5.2 Algoritam interpretiranja asocijativnih pravila sa TSK fazi pravilom

Postupak preklapanja posledičnih delova asocijativnih i TSK pravila se odvija na isti način kao i kod uslovnih delova, s tom razlikom što u posledičnim delovima pravila figuriše samo jedna varijabla *Posledice*. Pseudo-kod ovog algoritma je prikazan u Tabeli 4.3. Preseci egzaktnih vrednosti posledica asocijativnih pravila i TSK fazi pravila se smeštaju u vektor *ConsOverlap*.

U okviru algoritma za povezivanje i interpretiranje asocijativnih pravila preko TSK fazi pravila proverava se prvo nivo preklapanja posledičnih delova asocijativnih i TSK fazi pravila. U tu svrhu se koristi opisana promenljiva *ConsOverlap*. Nakon sortiranja po opadajućoj vrednosti, odbacuju se parovi asocijativnih i TSK pravila koji ne ispunjavaju određeni prag.

Za parove asocijativnih i TSK pravila koje zadovoljavaju prag vrednosti posledičnih delova pravila se nakon nalaženja odgovarajućeg para pravila množe

kumulativne vrednosti preseka uslovnih delova asocijativnog i TSK fazi pravila sa odgovarajućim presekom posledičnih delova pravila. Rezultat ove operacije je vektor *Similarity* koji predstavlja nivo preklapanja asocijativnog pravila i TSK fazi pravila. Prvi element sortiranih vrednosti vektora *Similarity* ima značenje najslabijeg asocijativnog i TSK fazi pravila.

Tabela 4.3 Algoritam interpretiranja asocijativnih pravila sa TSK fazi pravilom

---

```

Get vector ConsOverlap; vector CumulCausesOverlap;
      vector IndicesAssociate_TSKrules
Init IndicesPairs,k=0
SortedConsOverlap=sort(ConsOverlap)
while(SortedConsOverlap(i)>Threshold)
  k=k+1
  IndicesPairs(k)=IndicesAssociate_TSKrules(i)
end while
SortedCumulCausesOverlap=sort(CumulCausesOverlap)
k=1
while(SortedCumulCausesOverlap(k)!=EndOfVector)
  if(IndicesAssociate_TSKrules(k) foundIn(IndicesPairs))
    Similarity(k)=SortedCumulCausesOverlap(k)
    foundConsOverlap=find(ConsOverlap)
    with_sameIndices(SortedCumulCausesOverlap(k))
    Similarity(k)=Similarity(k)·foundConsOverlap
  end if
end while

```

---

### 4.5.3 Uprošćeni postupak određivanja sličnosti između asocijativnih i TSK fazi pravila

Tabela 4.4 prikazuje pojednostavljeni algoritam za određivanje sličnosti između asocijativnih i TSK fazi pravila. Kumulativno preklapanje između uslovnih delova svih TSK fazi pravila i svakog pojedinačnog asocijativnog pravila se usrednjavaju (dobijena suma se deli sa brojem TSK fazi pravila), a potom se sortiraju po opadajućim vrednostima. Što je veća dobijena vrednost za pojedino asocijativno pravilo, tim je pokrivenost asocijativnog pravila sa TSK fazi modelom veća.

Tabela 4.4 Algoritam određivanja sličnosti između asocijativnih i TSK fazi pravila

---

```

Get vector CumulCausesOverlap

```

---

---

```
for i=1 to number of association rules
  AverageCumulCausesOverlap(i)=
    sumAll(CumulCausesOverlap)/number of TSK rules
end for
sort(AverageCumulCausesOverlap,descending)
```

---



## 5. Rezultati selekcije obeležja

U ovom poglavlju su dati rezultati i analiza dva postupka selekcije obeležja korišćene baze podataka saobraćajnih udesa sa učešćem pešaka. Pre izlaganja rezultata selekcije obeležja postupcima kros-korelacije i PSO+ANN algoritama, dat je opis korišćene baze podataka.

### 5.1 Opis baze podataka

Baza podataka saobraćajnih udesa sa učešćem pešaka koja je korišćena sadrži podatke o saobraćajnim nezgodama u periodu od 2008. do 2011. godine za urbano područje Novog Sada. Inicijalni skup podataka sadrži ukupno 42 karakteristike i 752 saobraćajne nezgode, od kojih je zabeleženo 798 slučajeva sa nastradalim pešacima [100]. Korišćena baza podataka sadrži odvojene zapise podataka za obe uloge učesnika u saobraćajnim nezgodama: pešake i vozače. Podaci sadrže zvanične statističke zapise o saobraćajnim nezgodama (osnovni podaci o npr. lokaciji i vremenu udesa), učesnicima saobraćajnih nezgoda (poput njihove uloge, pola, starosti, itd.), vozilima (tipu vozila, godini proizvodnje, itd.) i sličnim informacijama. Svakodnevni policijski izveštaji su upotpunili opis pojedinih saobraćajnih udesa i obezbedili detaljniji opis pojedinih saobraćajnih nezgoda, npr. informacijama o načinu na koji su se učesnici kretali u saobraćajnim nezgodama ili detaljniji opis lokacije saobraćajne nesreće. Identifikovani su ključni faktori koji su imali uticaja na pojavu saobraćajnih nezgoda i grupisani su u četiri kategorije:

- vremenske karakteristike (informacije o vremenskim uslovima),
- karakteristike udesa,
- karakteristike puta (geometrija puta i njegovog okruženja) i
- karakteristike učesnika (personalni podaci).

Liste varijabli koje pripadaju ovim kategorijama su dati u narednim tabelama (Tabele 5.1 – 5.4). U svakoj tabeli je pored naziva varijable dato njeno značenje i kodovane vrednosti (kod) sa objašnjenjima. Izlazni atribut, tj. posledice saobraćajne nezgode (varijabla *Posledice*), sadrži četiri nivoa intenziteta (eng. *severity*) posledica saobraćajnih nezgoda: sa fatalnim ishodom, teško povređenim, lako povređenim i bez povređenih.

Tabela 5.1 Objašnjenja za grupu promenljivih: Vremenske karakteristike

Naziv	Značenje	Kod	Objašnjenje koda
<b>Atmosferske_prilike</b>	Atmosferske prilike (vremenski uslovi) na osnovu stanja kolovoza		Suva - čista
		1	Suva - sa nanosom prašine, peska i dr.
		2	Mokra - čista
		3	Mokra - zaprljana blatom, muljem i dr.
		4	Poledica - posuta so ili rizla
		5	Poledica - nije posuta
		6	Pokrivena snegom - razgrnut i posut
		7	Pokrivena snegom - nerazgrnut
<b>Vidljivost</b>	Svetlosne prilike	1	Dobra, dnevna
		3	Slaba, dnevna
		4	Noćna, nedovoljno osvetljen put
		5	Noćna, neosvetljen put
<b>Dan_u_nedelji</b>	Dan u nedelji	1	Ponedeljak
		2	Utorak
		3	Sreda
		4	Četvrtak
		5	Petak
		6	Subota
		7	Nedelja
<b>Godina</b>	Godina kada se nesreća desila	2008	2008. godina
		2009	2009. godina
		2010	2010. godina
		2011	2011. godina
<b>Godisnja_doba</b>	Doba godine	1	Proleće
		2	Leto
		3	Jesen
		4	Zima
<b>Doba_dana</b>	Vremenski opseg u toku dana	1	0-5 h
		2	6-11 h
		3	12-17 h
		4	18-23 h
<b>Radni_dani_vikend</b>	Radni dan ili vikend	1	Radni dani
		2	Vikend

Tabela 5.2 Objašnjenja za grupu promenljivih: Karakteristike udesa

Naziv	Značenje	Kod	Objašnjenje koda
<b>Kontakt_pesak_vozilo</b>	Mesto kontakta između vozila i pešaka	1	Prednjim čeonim
		2	Prednjim levim
		3	Prednjim desnim
		4	Bočnim levim delom
		5	Bočnim desnim delom
		7	Zadnjim levim delom
		8	Zadnjim desnim delom
		9	Zadnjim delom
		0	Ostalo
		<b>Vid_SN_vrsta</b>	Vrsta saobraćajne nesreće (SN)
<b>Vrsta_nezgode_prema_posledicama</b>	Vrsta SN prema posledicama	1	SN sa pogunulim
		2	SN sa povređenim
		3	SN sa materijalnom štetom
<b>Glavni_uzrok</b>		101-353	Glavni uzrok SN
<b>NN_nezgoda</b>	Učesnik (ni)je napustio mesto SN	1	Da
		2	Ne
<b>Svojtvo_lica</b>	Uloga učesnika u SN	1	Vozač
		2	Pešak
<b>Promena_vida_prevoza</b>	Promena vida prevoza	1	Izlazak iz PA
		2	Izlazak iz TV
		3	Izlazak iz BUS
		0	Bez promene
<b>Posledice</b>	Posledice SN	0	Lice nepovređeno
		1	Smrt na licu mesta
		2	Smrt za vreme prevoza do bolnice
		3	Smrt do 30 dana
		4	Teške telesne povrede
5	Lake telesne povrede		
<b>Posledice_grupisano</b>	Grupisane posledice	1	Poginuli
		2	Teško povređeni
		3	Lako povređeni
		4	Materijalna šteta

Tabela 5.3 Objašnjenja za grupu promenljivih: Karakteristike puta

Naziv	Značenje	Kod	Objašnjenje koda
<b>Karakteristike_puta</b>	Karakteristike puta	1	Prav-sa suženjem kolovoza na mestu SN
		2	Prav-sužen kolovoz celom deonicom
		3	Prav-bez suženja
		4	Krivina-pregledna
		5	Krivina-nepregledna
		6	Prevoj-pregledan
		7	Prevoj-nepregledan
		8	Neregulisana raskrsnica-pregledna
		9	Neregulisana raskrsnica-nepregledna
		10	Regulisana raskrsnica-saobraćajnim znakovima
		11	Regulisana raskrsnica-semaforima
<b>Osobine_povrsine_kolovoza</b>	Osobine površine kolovoza	1	Asfalt-gladak, ravan i bez oštećenja
		2	Asfalt-gladak, istalasan i sa oštećenjima
		3	Asfalt-rapav, ravan i bez oštećenja
		4	Asfalt-istalasan i sa oštećenjima
		5	Beton-ravan i bez oštećenja
		6	Beton-neravan i sa oštećenjima
		7	Kamena kocka ili kaldrma-ravna, bez oštećenja
<b>Stanje_kolovoza</b>	Stanje kolovoza	1	Suva-čista
		2	Suva-sa nanosom prašine, peska i dr.
		3	Mokra-čista
		4	Mokra-zaprljana blatom, muljem i dr.
		5	Poledica-posuta so ili rizla
		6	Poledica-nije posuta
		7	Pokrivena snegom-razgrnut i posut
		8	Pokrivena snegom-nerazgrnut
<b>Ulica_1</b>	Lokacija SN	20-7010	Kod ulice
<b>Kilometar_puta_broj_ulice_1</b>	Lokacija SN	-	Pozicija u ulici
<b>Ulica_2</b>	Lokacija SN	0-7010	Kod ulice
<b>Kilometar_puta_broj_ulice_2</b>	Lokacija SN	-	Pozicija u ulici
<b>Raskrsnica</b>	Jeste/nije	1	Da
	raskrsnica	2	Ne
<b>SM_kretanja_voz</b>	Smer kretanja	1	Pravo

*Rezultati selekcije obeležja*

	vozila	2	Levo
		3	Desno
		4	Ostalo
		5	Unazad
<b>P_prelaz</b>	Pešački prelaz	1	Na pešačkom prelazu
		2	Van pešačkog prelaza
		3	Trotoar
		4	Ostalo
<b>Prelazenje_kol</b>	Prelaženje kolovoza	1	Sa leve na desnu stranu
		2	Sa desne na levu stranu
		3	Ostalo

Tabela 5.4 Objašnjenja za grupu promenljivih: Karakteristike učesnika (personalni podaci)

Naziv	Značenje	Kod	Objašnjenje koda
<b>Godina_rodjenja</b>	Godina rođenja učesnika u SN	-	Godina rođenja učesnika
<b>Pol</b>	Pol učesnika u SN	0-4	Muški
		5-9	Ženski
<b>Kolicina_alkohola</b>	Količina alkohola u organizmu	-	Količina alkohola [‰]
<b>Pojavni_oblik_ponasanja</b>	Pojavni oblik ponašanja		Prolazak kroz crveno svetlo
		1	Pretrčavanje kolovoza
		2	Nepropisno kretanje unazad
		3	Neprikladna brzina
		4	Neustupanje prvenstva
		5	prolaza pesacima
		6	Kretanje u suprotnom smeru
		7	Vožnja pod uticajem narkotika
		0	Ostalo
<b>Pojavni_oblik_ponasanja_2</b>	Pojavni oblik ponašanja	1	Pretrčavanje kolovoza
		2	Nepropisno kretanje unazad
		3	Neprikladna brzina
		4	Neustupanje prvenstva
		5	prolaza pešacima
		6	Kretanje u suprotnom smeru
		7	Vožnja pod uticajem narkotika
<b>ID_vozila_lica</b>	ID vozila lica	1-278	
<b>Kategorija_vozacke_dozvole</b>	Kategorija vozacke dozvole	A	Kategorija A-motocikl
		B	Kategorija B-motorna

		C	vozila
		D	Kategorija C-m. vozila za prevoz tereta Kategorija D-m. vozila za prevoz putnika
<b>Godina_polaganja_vozackog_ispita</b>	Godina polaganja vozačkog ispita	-	Godina polaganja vozačkog ispita
<b>Redni_broj_vozila</b>	Redni broj vozila	1	Prvo vozilo po redosledu navođenja u SN
		2	Drugo vozilo po redosl. navođenja u SN
<b>Vrsta_vozila</b>	Vrsta vozila	1	Bicikl
		2	Bicikl sa motorom, radna mašina
		3	Motocikl
		4	Traktor
		5	Putnički automobil
		6	Kombi vozilo
		7	Autobus-trolejbus
		8	Teretno vozilo
		9	Specijalno i radno vozilo
		10	Vučno vozilo
		11	Priključno vozilo
		12	Zaprežno vozilo
		13	Železičko vozilo
		14	Tramvaj
<b>Zemlja_registracije_vozila</b>	Zemlja registracije vozila	100-822	Registarska oznaka vozila države
<b>Godina_proizvodnje_vozila</b>	Godina proizvodnje vozila	-	Starost vozila (godište)
<b>Pol_grupisano</b>	Pol učesnika u SN (grupisano)	1	Muški
		2	Ženski
		0	Nepoznato
<b>Godine_rodjenja_ok</b>	Godine rođenja ok	-	Godine rođenja učesnika
<b>Starost</b>	Godine učesnika u SN	-	Godine učesnika proračunate na osnovu godine rođenja

		1	0-14
		2	15-29
<b>Starosne_grupe</b>	Starosne grupe	3	30-44
		4	45-64
		5	65>
		1	PA
		2	TV
<b>Vrsta_vozila_grupisano</b>	Vrsta vozila grupisano	3	BUS
		4	Motocikl
		5	Ostalo

## 5.2 Selekcija obeležja kros-korelacijom

Nakon postupka filtriranja podataka iz date baze podataka o saobraćajnim udesima sa učešćem pešaka sa 42 varijable, odnosno isključivanja redundantnih i nepotpunih podataka, preostao je skup podataka od 25 varijable čija je lista (indeks i naziv) data u Tabeli 5.5.

Tabela 5.5 Polazne varijable filtriranog skupa podataka o saobraćajnim udesima sa učešćem pešaka

Redni broj	Naziv varijable
1	Atmosferske prilike
2	Vidljivost
3	Kontakt pešak vozilo
4	Karakteristike puta
5	Osobine površine kolovoza
6	Stanje kolovoza
7	Dan u nedelji
8	Količina alkohola
9	Pojavni oblik ponašanja
10	Pojavni oblik ponašanja 2
11	Promena vida prevoza
12	Vrsta vozila
13	Starost vozila
14	Pol grupisano
15	Starost pešaka
16	Godišnja doba



17	Doba dana
18	Radni dani vikend
19	Starosne grupe
20	Raskrsnica
21	Vrsta vozila grupisano
22	Smer kretanja vozila
23	Pešački prelaz
24	Prelaženje kolovoza
25	Posledice grupisano

Jednostavan uvid u međusobne odnose između promenljivih omogućava kros-korelaciona analiza. Matrica kros-korelacija za posmatranih 25 varijabli, gde je poslednja varijabla *Posledice\_groupisano*, je data na Slici 5.1. Posmatrajući elemente matrice sa najvećim vrednostima (+1, -1), odnosno najvećim intezitetom boja (plavo, crveno), može se zaključiti da najjače korelacije postoje između sledećih parova varijabli:

Raskrsnica-Karakteristike puta -0.99

Starosne grupe-Starost pešaka 0.98

Radni dani, vikend-Dan u nedelji 0.70

Atmosferske prilike-Stanje kolovoza 0.61

Kontakt pešak, vozilo-Smer kretanja vozila 0.59

Pešački prelaz-Karakteristike puta -0.51

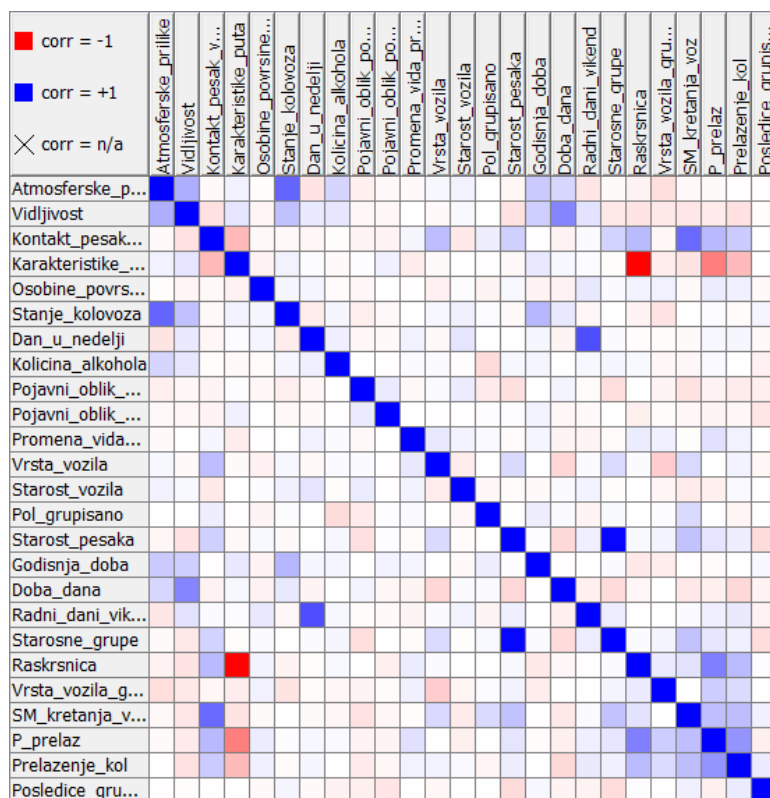
Raskrsnica-Pešački prelaz 0.50

Vidljivost-Doba dana 0.48

Karakteristike puta-Kontakt pešak, vozilo -0.28

Karakteristike puta-Prelaženje kolovoza -0.27

## Povećanje razumljivosti modela zasnovanog na pravilima



Slika 5.1 Matrica kros-korelacija za posmatranih 25 varijabli

Jasno je da parovi varijabli *Raskrsnica-Karakteristike puta* i *Starosne grupe-Starost pešaka* predstavljaju varijable koje su gotovo identične, jer su korelacije između njih veće od 0.98. Takve varijable ne treba uzimati istovremeno kao ulazne varijable prilikom generisanja predikcionog ili klasifikacionog modela.

Parovi varijabli *Radni dani, vikend-Dan u nedelji* i *Atmosferske prilike-Stanje kolovoza* imaju značajnu korelaciju, ali se ipak ne radi o identičnim varijablama. Ove varijable, između kojih su korelacije u opsegu 0.6-0.7, nisu identične, međutim zbog sličnosti je preporuka da se ne uzimaju u razmatranje istovremeno za ulaz u predikcioni ili klasifikacioni model.

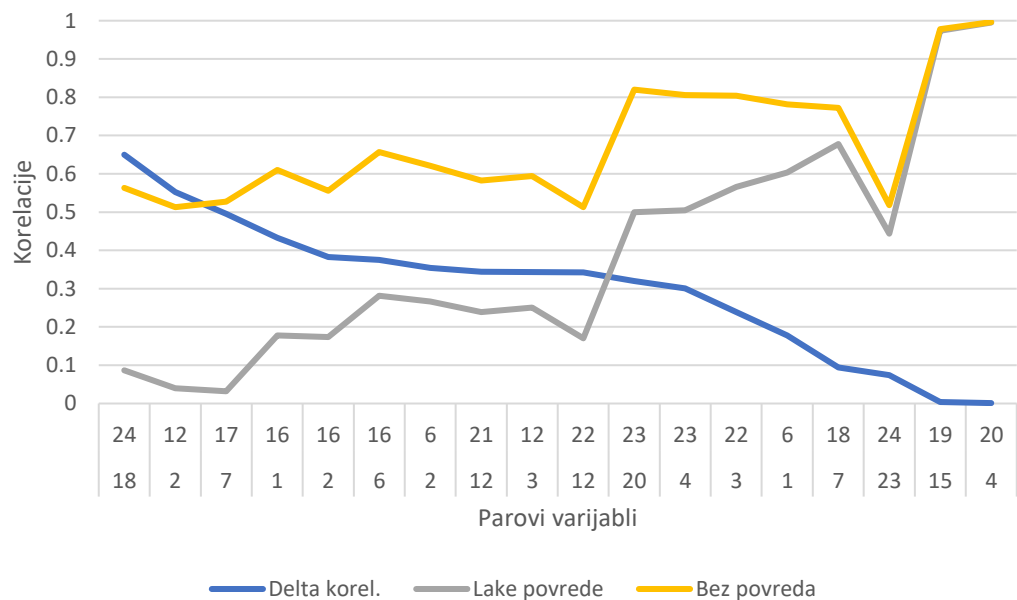
Najinteresantniji deo kros-korelacione matrice je njena poslednja kolona/vrsta, jer se tu vidi veza između posledične varijable *Posledice grupisano* i ostalih varijabli. Uočljiva je vrlo slaba zavisnost posledične varijable sa ostalim varijablama. Najveću korelaciju pokazuje varijabla *Starosne grupe* (-0.135), a potom *Prelazanje kolovoza* (0.08), *Pojavni oblik ponašanja* (-0.10) i *Pešački prelaz* (-0.06).

U nastavku analize i selekcije obeležja polaznog skupa podataka sa 25 varijabli je primenjen postupak Razlike parcijalnih kros-korelacionih matrica, izložen u poglavlju 4.1. Ako se posmatra posledična varijabla *Posledice grupisano*, za nju je

karakteristično da ima četiri moguće vrednosti (“smrtni slučajevi”, “teške povrede”, “lake povrede”, “bez povreda”). Kada se podeli skup uzoraka iz baze podataka na četiri grupe koje imaju iste vrednosti varijable *Posledice grupisano*, dobiju se skupovi sa po 13, 188, 577 i 20 uzoraka, respektivno. U daljem tekstu sledi analiza uticaja među varijablama posmatranjem parova skupova uzoraka koji odgovaraju sledećim vrednostima varijable *Posledice*:

“lako povređeni” – “bez povreda”, “teško povređeni” - “bez povreda”, “teško povređeni” - “lako povređeni”.

Interesantno je analizirati varijable nakon proračunavanja prvog slučaja, tj. skupova uzoraka sa vrednošću varijable *Posledice* “lake povrede” i uzoraka sa vrednošću varijable *Posledice* “bez povreda”. Na osnovu proračuna kros-korelacionih matrica i njihove razlike za ova dva skupa uzoraka, dobijaju se vrednosti prikazane u na Slici 5.2.



Slika 5.2 Kros-korelacije za skupove podataka sa vrednostima “lake povrede” i “bez povreda”

Cilj analize je da se identifikuju varijable koje su značajno promenile svoj međusobni uticaj u okviru skupova uzoraka koji odgovaraju udesima sa lakim povredama i udesima bez povreda. Da bi se takvi slučajevi detektovali, neophodno je posmatrati istovremeno dva uslova: visoke vrednosti kros-korelacija u jednom od skupova i razlike kros-korelacija ta dva skupa (veličina *Delta korel*). Ako se usvoji prag od 0.2 za *Delta korel* i 0.6 kao prag za pojedinačne kros-korelacije, ove uslove

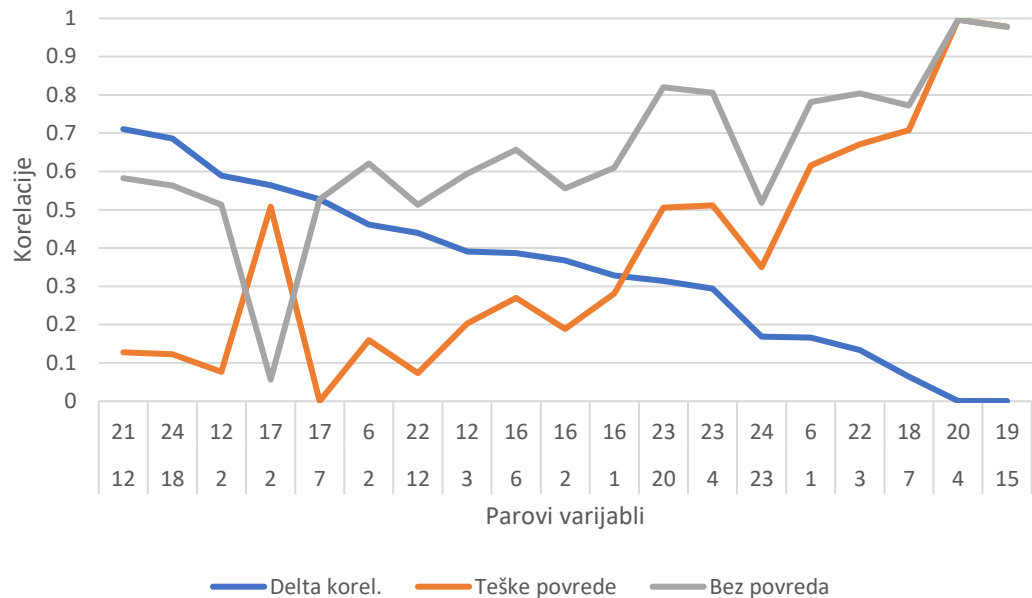
zadovoljava nekoliko parova varijabli (videti Sliku 5.2). Apscisa na Slici 5.2 sadrži indekse izdvojenih parova varijabli dva skupa uzoraka koja su razmatrana. U Tabeli 5.6 je data lista selektovanih parova varijabli sa najvećim promenama kros-korelacija za slučajeve udesa sa lakim povredama i bez povreda.

Tabela 5.6 Lista selektovanih parova varijabli sa najvećim promenama kros-korelacija: slučajevi sa lakim povredama i bez povreda

<b>Parovi varijabli</b>	<b>Delta korelacija</b>	<b>Kros-korelacija (slučajevi sa lakim povredama)</b>	<b>Kros-korelacija (slučajevi bez povreda)</b>
Radni dani vikend-Prelaženje kolovoza (18-24)	0.6860	0.1225	0.5634
Atmosferske prilike-Godišnja doba (1-16)	0.43254	0.17743	0.60997
Stanje kolovoza-Godišnja doba (6-16)	0.37501	0.28159	0.6566
Vidljivost-Stanje kolovoza (2-6)	0.35429	0.26639	0.62068
Raskrsnica-Pešacki prelaz (20-23)	0.31965	0.50013	0.81978
Karakteristike puta-Pešacki prelaz (4-23)	0.30113	-0.50474	-0.80587
Kontakt pešak vozilo-Smer kretanja vozila (3-22)	0.23898	0.56543	0.80441

U slučaju varijabli *Radni dani vikend-Prelaženje kolovoza (18-24)* je promena kros-korelacija (*Delta korel*) najveća, međutim najveća kros-korelacija (slučaj bez povreda) je svega 0.56, za čega se ne može smatrati da postoji visoka povezanost među varijablama. S druge strane, kod poslednja tri para varijabli iz Tabele 5.6, tj. *Raskrsnica-Pešacki prelaz*, *Karakteristike puta-Pešacki prelaz*, *Kontakt pesak vozilo-Smer kretanja vozila (20-23, 4-23, 3-22)*, postoje visoke kros-korelacije za slučaj vrednosti *Posledice* „bez povreda“, uz istovremeno postojanje značajnog pada korelacije u slučaju vrednosti „lake povrede“. Ovo može biti indikacija postojanja potencijalnog uticaja ovih varijabli na težinu posledica saobraćajnih udesa.

Postupak analize je ponovljen nad podacima za slučaj skupa uzoraka sa vrednošću varijable *Posledice* “teške povrede” i skupa uzoraka sa vrednošću varijable *Posledice* “bez povreda”. Na osnovu proračuna kros-korelacionih matrica i njihove razlike za ova dva skupa uzoraka, dobijaju se vrednosti prikazane u na Slici 5.3.



Slika 5.3 Kros-korelacije za skupove podataka sa vrednostima “teške povrede” i “bez povreda”

Potrebno je da se identifikuju varijable koje su značajno promenile svoj međusobni uticaj u okviru skupova uzoraka koji odgovaraju udesima sa teškim povredama i udesima bez povreda. Da bi se takvi slučajevi otkrili, neophodno je istovremeno posmatrati dva uslova: visoke vrednosti kros-korelacija u jednom od skupova i razlike kros-korelacija ta dva skupa (*Delta korel*). Za iste pragove od 0.2 i 0.6 za *Delta korel* i za pojedinačne kros-korelacije, ove uslove su zadovoljili parovi varijabli dati u Tabeli 5.7.

Tabela 5.7 Lista selektovanih parova varijabli sa najvećim promenama kros-korelacija: skupovi uzoraka sa vrednostima varijable Posledice: “teške povrede” i “bez povreda”

Parovi varijabli	Delta korelacija	Kros-korelacija (slučajevi sa teškim povredama)	Kros-korelacija (slučajevi bez povreda)
Vidljivost-Stanje kolovoza (2-6)	0.46149	0.15919	0.62068
Kontakt pešak vozilo-Vrsta vozila (3-12)	0.3912	0.20299	0.59419
Stanje kolovoza-Godišnja doba (6-16)	0.38719	0.26941	0.6566

Atmosferske prilike-Godišnja doba (1-16)	0.32836	0.28161	0.60997
Raskrsnica-Pešacki prelaz (20-23)	0.31436	0.50542	0.81978
Karakteristike puta-Pešacki prelaz (4-23)	0.29414	-0.51173	-0.80587

U slučaju varijabli *Vidljivost-Stanje kolovoza (2-6)* je promena kros-korelacija (*Delta korel*) najveća, a najveća kros-korelacija (slučaj bez povreda) je 0.62, tako da se može smatrati da postoji potencijalni uticaj ovih varijabli na težinu posledica udesa. Pored toga, kod poslednja dva para varijabli iz Tabele 5.7, tj. *Raskrsnica-Pešacki prelaz*, *Karakteristike puta-Pešacki prelaz (20-23, 4-23)*, postoje visoke kros-korelacije za slučaj „bez povreda“, uz istovremeno postojanje značajnog pada korelacije u slučaju težih povreda. Ovo može biti indikacija postojanja potencijalnog uticaja ovih varijabli na težinu posledica udesa.

Parovi varijabli *Raskrsnica-Pešacki prelaz* i *Karakteristike puta-Pešacki prelaz (20-23 i 4-23)* imaju uniforman obrazac vrednosti za slučaj kada su posledice bez povreda, na šta ukazuju visoke korelacije date u Tabelama 5.6 i 5.7. Isti parovi varijabli imaju značajno manje korelacije, tj. zavisnost za slučaj kada posledice odgovaraju povredama (lakim i teškim). Promena u zavisnosti između varijabli u slučaju različitih posledica ukazuje da postoji kombinovani, nelinearni uticaj ovih varijabli na posledice. Interesantno je napomenuti da se parovi varijabli *Stanje kolovoza-Godišnja doba* i *Atmosferske prilike-Godišnja doba (6-16 i 1-16)* pojavljuju i imaju značajne promene u oba posmatrana slučaja.

U skupovima uzoraka koji odgovaraju slučajevima teških i lakih povreda nije primećena varijacija u promeni zavisnosti između varijabli. Daleko najveća promena je uočena na paru *Kontakt pešak vozilo-Smer kretanja vozila (3-22)* od svega 0.11.

### **5.3 Rezultati selekcije pomoću PSO+ANN algoritma**

U cilju nalaženja najpogodnijih podskupova varijabli, koje mogu poslužiti za generisanje predikcionih modela težine udesa, korišćen je algoritam PSO+ANN naveden u poglavlju 4.1. Naime, PSO algoritam se koristi kao postupak pretrage prostora mogućih podskupova ulaznih varijabli, dok ANN služi za evaluaciju fitnes funkcije svih razmatranih podskupova varijabli. Razmatrani su slučajevi za različit broj ulaznih varijabli: od dva do pet ulaza. Polazni skup od 25 ulaznih veličina odgovara listi varijabli datih na ordinati Slike 5.4 i on je sličan listi varijabli datih u

Tabeli 5.5 koja je korišćena kod postupka kros-korelacione analize. U metod pretrage *PSO* algoritmom ugrađena je neuronska mreža sa jednim skrivenim slojem i sa kriterijumskom funkcijom u obliku srednje apsolutne greške između stvarnog i estimiranog odziva. Za slučaj sa pet ulaznih varijabli, neuronska mreža ima jedanaest neurona na skrivenom sloju i jedan izlazni neuron. Izlazni neuron ima značenje variable *Posledice grupisano*.

Rangiranje (selekcija) važnosti obeležja na osnovu *PSO+ANN* metode sa 100 jedinki (rešenja) i 50 iteracija. Kompletan postupak je ponovljen 20 puta, kako bi se usrednjila stohastička priroda metode pretrage. Usvojen je koeficijent inercije 0.754, a za personalni ( $c_1$ ) i globalni koeficijent učenja ( $c_2$ ) je usvojeno  $c_1=c_2=1.5457$ . Vrednosti najboljih jedinki, tj. jedinki sa najboljom kombinacijom varijabli su iskorišćene za formiranje funkcije rangiranja koja ima sledeći oblik:

$$\sum_i^m \sum_j^n \left(1 - \frac{c_{ij} - c_{min}}{c_{max} - c_{min}}\right) \quad (5.1)$$

$m$  – broj ponavljanja

$n$  – broj varijabli

$c_{ij}$  – vrednost rešenja  $i$ -tog ponavljanja u kome je učestvovala  $j$ -ta varijabla

$c_{max}$  – najslabiji postignuti rezultat

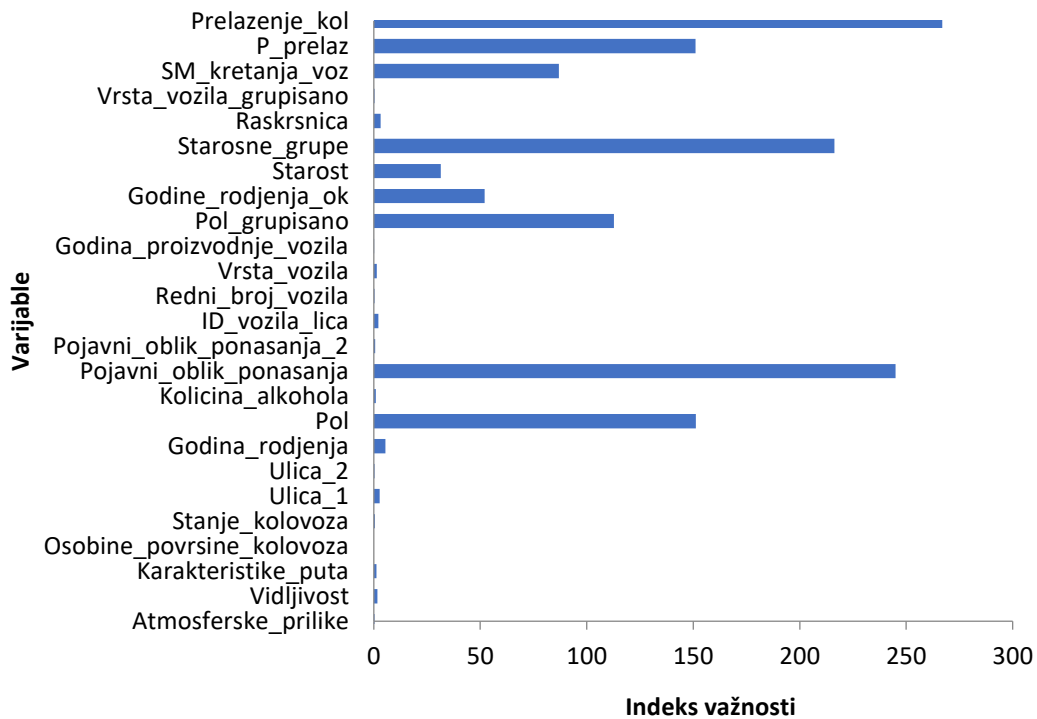
$c_{min}$  – najbolji postignuti rezultat

Za slučaj kada je razmatran izbor najboljih pet obeležja, *PSO+ANN* algoritam je selektovao varijable prikazane na Slici 5.4. Stubići iznad broja odgovarajuće varijable predstavljaju indeks važnosti varijable izračunate po formuli (5.1). Na osnovu indeksa važnosti iz formule (5.1), može se odabrati podskup varijabli [9, 11, 20, 24, 25], odnosno varijable [*Pol*, *Pojavni oblik ponašanja*, *Starosne grupe*, *Pesački prelaz*, *Prelaženje kolovoza*]. Važno je napomenuti da je u toku rada algoritma *PSO+ANN* postignuta kombinacija obeležja [11, 17, 20, 24, 25], odnosno varijabli [*Pojavni oblik ponašanja*, *Pol grupisano*, *Starosne grupe*, *Pesački prelaz*, *Prelaženje kolovoza*], sa minimalnom vrednošću funkcije cilja 0.026.

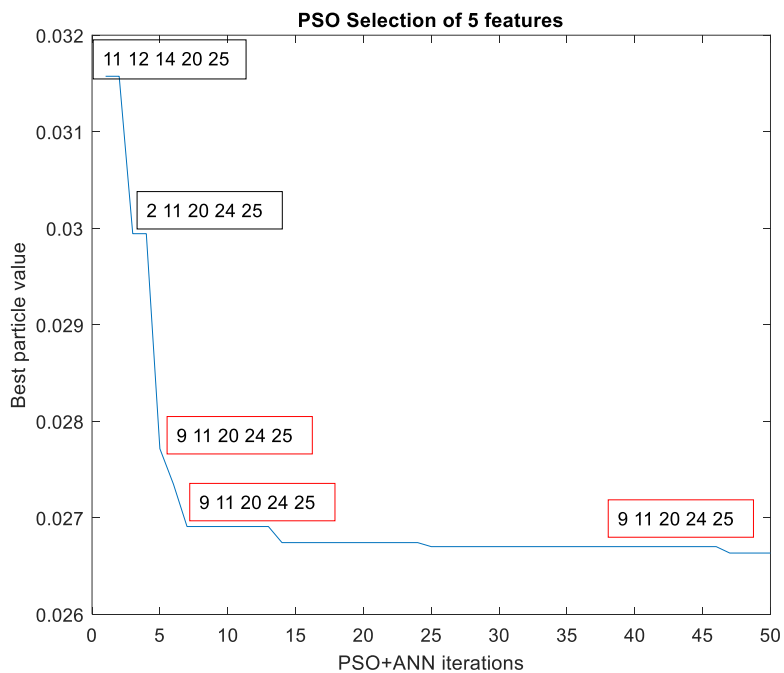
Kao ilustracija konvergencije *PSO+ANN* algoritma, na Slici 5.5 prikazano je 50 iteracija jednog izvršavanja *PSO* algoritma (na slučajan način odabran jedan tok izvršavanja). Algoritam je u prvih nekoliko iteracija konvergirao ka najboljoj kombinaciji obeležja [9, 11, 20, 24, 25], odnosno varijable [*Pol*, *Pojavni oblik ponašanja*, *Starosne grupe*, *Pesački prelaz*, *Prelaženje kolovoza*], gde je nakon

## Povećanje razumljivosti modela zasnovanog na pravilima

sedme iteracije i dalje smanjivao funkciju cilja, sa istom kombinacijom obeležja. Postignuto rešenje predstavlja jedno od najboljih rešenja.



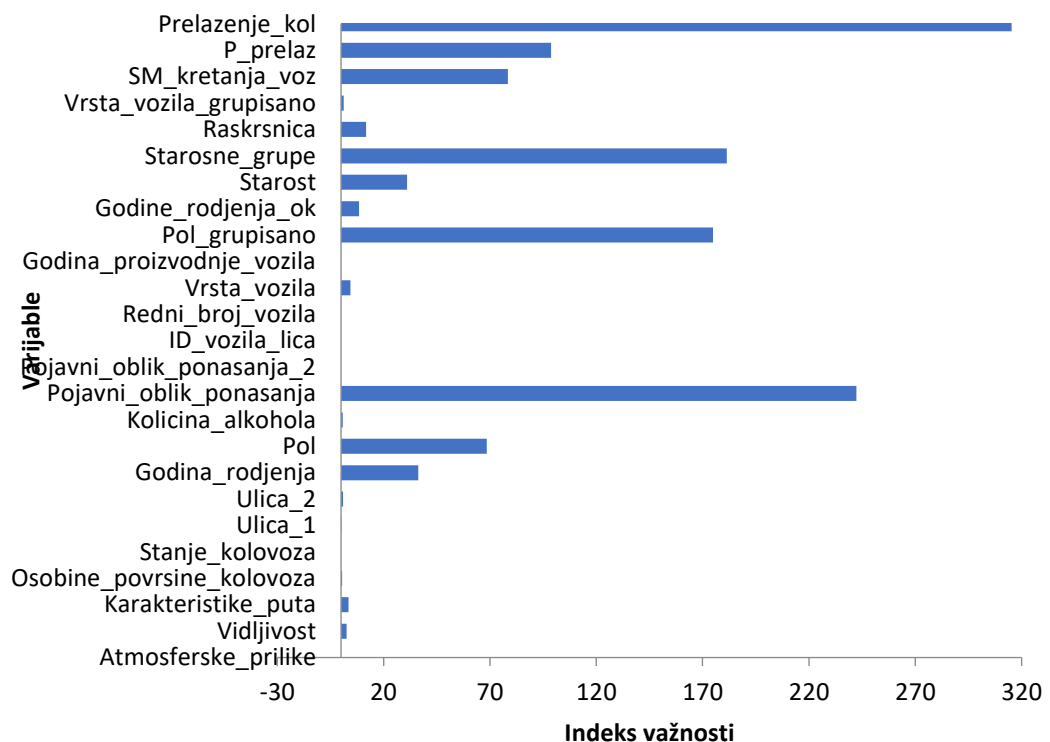
Slika 5.4 Izbor pet varijabli sa PSO+ANN algoritmom



Slika 5.5 Konvergencija PSO+ANN algoritma prilikom izbora pet varijabli

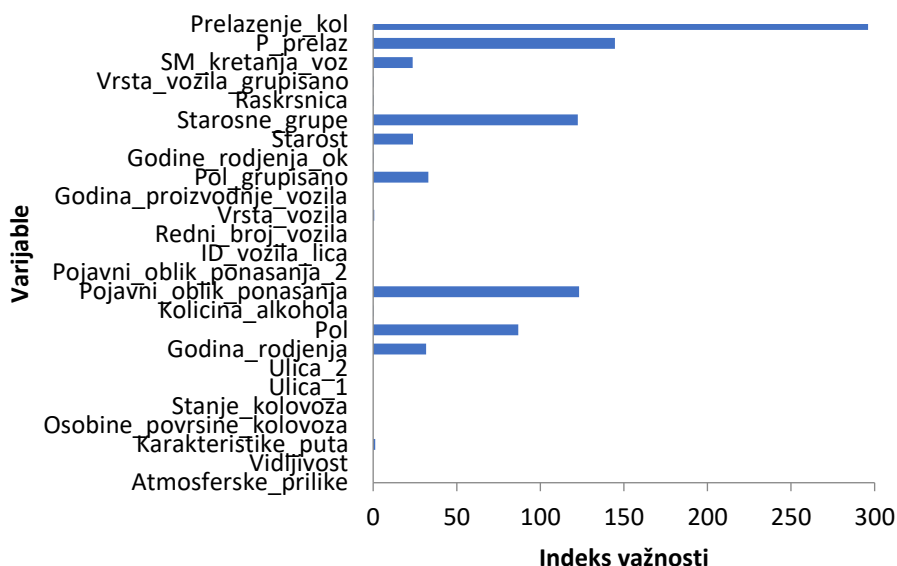


Za slučaj kada je razmatran izbor najbolje četiri ulazne varijable pomoću indeksa važnosti iz formule (5.1), [11, 17, 20, 25] se pokazalo kao najbolja kombinacija obeležja, odnosno varijable [*Pojavni oblik ponašanja, Pol\_grupisano, Starosne grupe, Prelaženje kolovoza*] se pokazalo kao najbolja kombinacija varijabli (Slika 5.6). Najbolja postignuta kombinacija varijabli u toku rada PSO+ANN algoritma je [11, 20, 24, 25], odnosno varijable [*Pojavni oblik ponašanja, Starosne grupe, Pesački prelaz, Prelaženje kolovoza*], sa vrednošću funkcije cilja 0.0280.



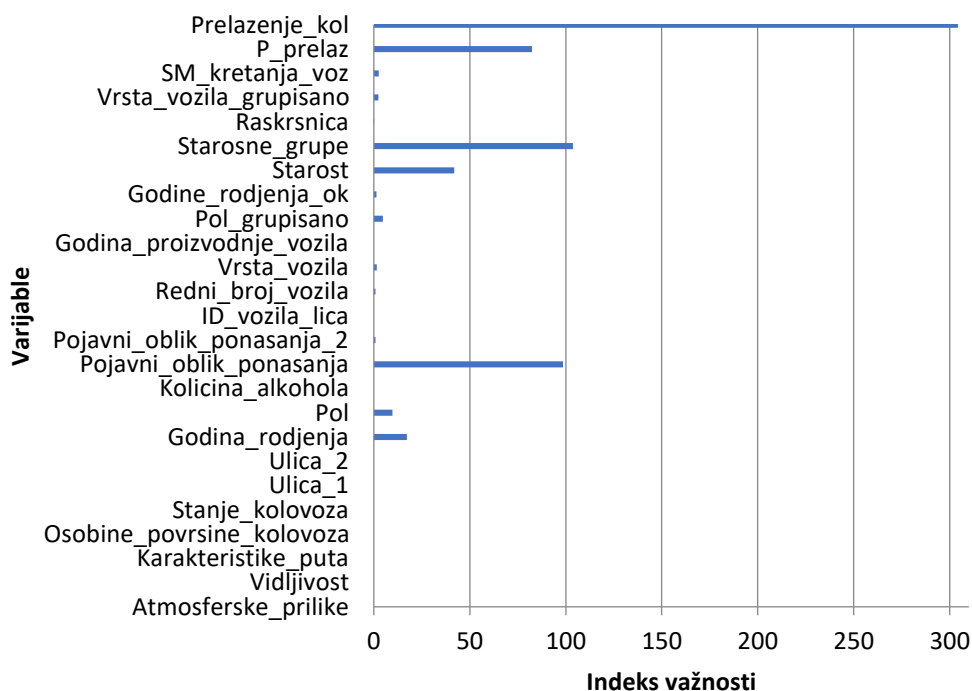
Slika 5.6 Izbor četiri varijable sa PSO+ANN algoritmom

Za slučaj kada je razmatran izbor najbolje tri varijable koristeći kriterijum indeksa važnosti (5.1), [11, 24, 25] se pokazalo kao najbolja kombinacija obeležja (Slika 5.7), odnosno varijable [*Pojavni oblik ponašanja, Pešački prelaz, Prelaženje kolovoza*]. Najbolja pojedinačna postignuta kombinacija u toku rada PSO+ANN algoritma je [11, 20, 25], tj. [*Pojavni oblik ponašanja, Starosne grupe, Prelaženje kolovoza*], sa vrednošću funkcije cilja 0.0306.



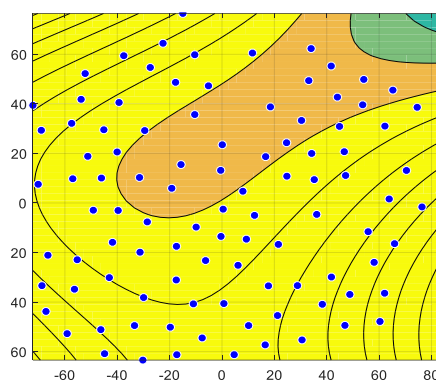
Slika 5.7 Izbor tri varijable sa PSO+ANN algoritmom

Za slučaj kada je razmatran izbor najbolje dve varijable po kriterijumu indeksa važnosti, [20, 25] se pokazalo kao najbolja kombinacija obeležja (Slika 5.8), tj. [Starosne grupe, Prelazanje kolovoza]. Najbolja pojedinačna postignuta kombinacija u toku rada PSO+ANN algoritma je [20, 25], tj. [Starosne grupe, Prelazanje kolovoza], sa vrednošću funkcije cilja 0.034804.

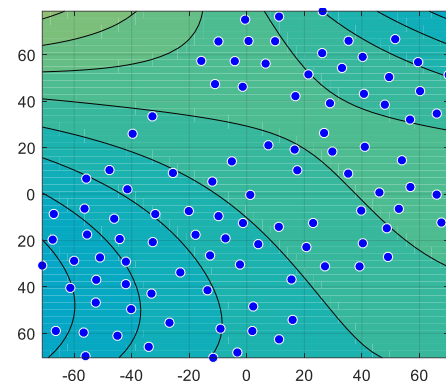


Slika 5.8 Izbor dve varijable sa PSO+ANN algoritmom

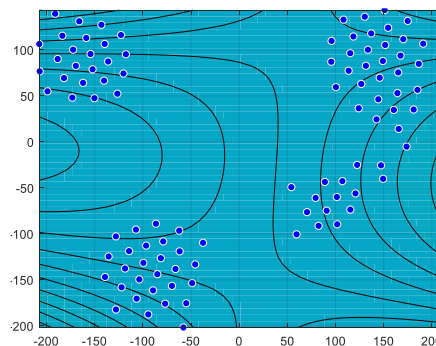
Na Slici 5.9 prikazan je raspored rešenja 100 jedinki nakon prve, pete, dvadesete i pedesete iteracije PSO+ANN algoritma. Da bi se raspored rešenja mogao prikazati, kombinacije petorke obeležja jedinki su transformisane u 2D reprezentaciju pomoću *tSNE* algoritma [101]. Ilustracija na Slici 5.9.a) pokazuje ocekivano veću distribuciju nakon prve iteracije, tj. raznovrsnost rešenja, dok se u kasnijim iteracijama primećuje da se disperzija rešenja postepeno smanjuje. Tačnije, rezultat je da u kasnijim iteracijama dolazi do grupisanja rešenja u oblastima sa nižim vrednostima fitnes funkcije PSO+ANN algoritma. Plavom bojom je na dijagramima označena niža vrednost, a žutom je prikazana najviša vrednost fitnes funkcija.



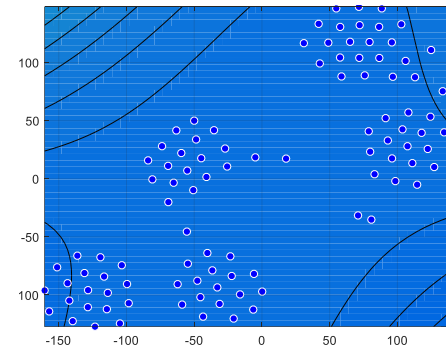
a) Pozicija rešenja nakon prve iteracije



b) Pozicija rešenja nakon pete iteracije



c) Pozicija rešenja nakon 20-te iteracije



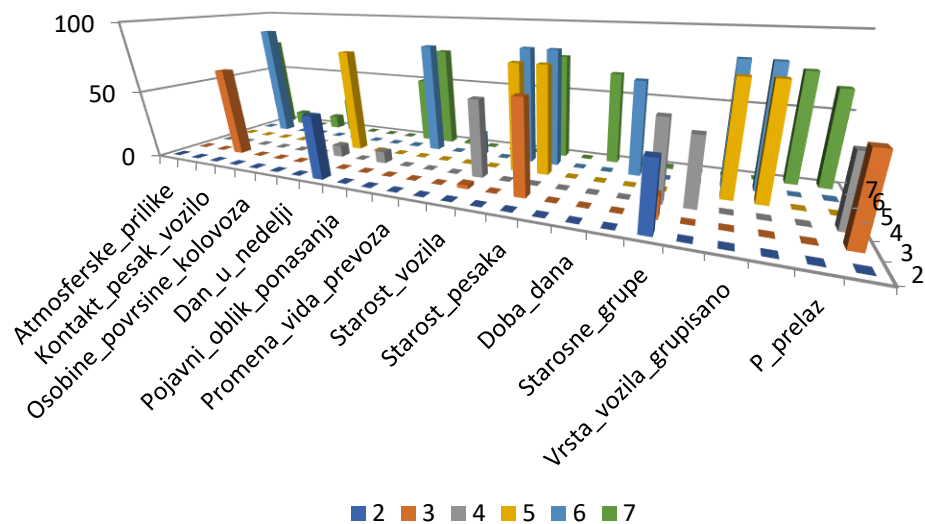
d) Pozicija rešenja nakon 50-te iteracije

Slika 5.9 Pozicije rešenja podskupova selektovanih varijabli tokom rada PSO+ANN algoritma pri izboru 5 varijabli

Postupak selekcije varijabli je ponovljen istim PSO+ANN algoritmom, ali uz promenu nekoliko parametara, odnosno povećan je broj iteracija na 100 i

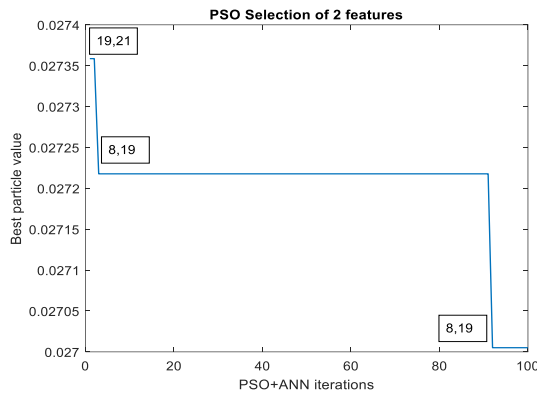
## Povećanje razumljivosti modela zasnovanog na pravilima

promenjen je globalni koeficijent učenja sa  $c_2=1.5457$  na  $c_2=0.85$ . Postupak je ponavljan 20 puta kao i u prethodnim slučajevima. U nastavku su dati rezultati za izvršavanje algoritma prilikom selekcije varijabli za interval od dve varijable do 7 varijabli. Na Slici 5.10 dat je indeks zastupljenosti svih varijabli u selektovanim podskupovima od dve do sedam varijabli. Varijable koje dominantno figurišu u selektovanim podskupovima su *Pol grupisano*, *Pojavni oblik ponašanja 2*, *Starosne grupe*, *Starost pešaka*, *Vrsta vozila grupisano*, *Smer kretanja vozila* i *Prelaženje kolovoza*.

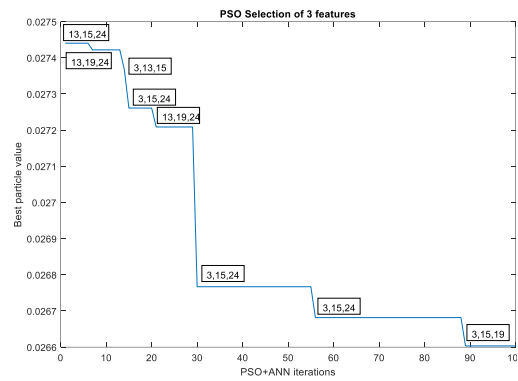


Slika 5.10 Zastupljenost varijabli za selektovani interval od 2 do 7 varijabli

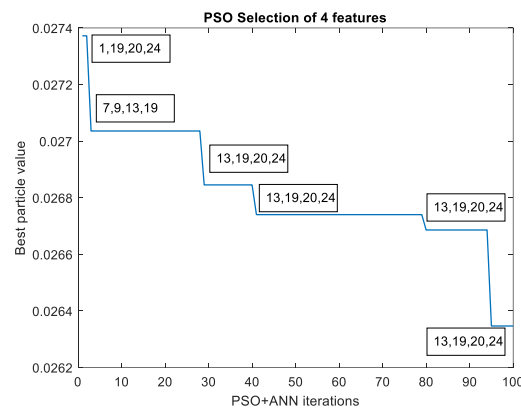
Na Slici 5.11 su prikazane iteracije za slučajeve izbora varijabli kada je tražen broj varijabli u intervalu od dve do sedam varijabli. Generalno, sa svih dijagrama je uočljiva pojava da se najbolje rešenje pronalazi u prvih dvadeset iteracija, što je delom posledica manje vrednosti globalnog koeficijenta učenja  $c_2$ . Pošto je personalni koeficijent učenja veći od globalnog, posledica je da algoritam pretrage više insistira na poboljšanju pojedinih rešenja, a ne na traženju njihovih alternativa.



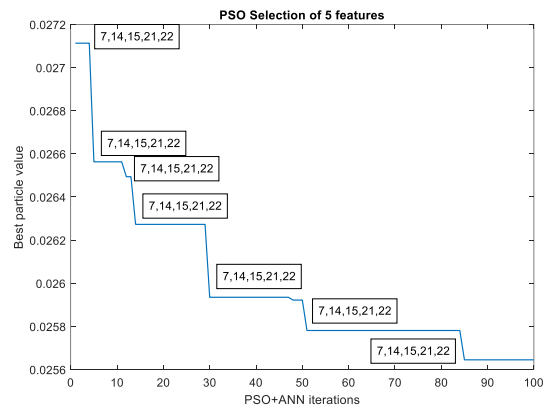
a) Tok selekcije 2 varijable



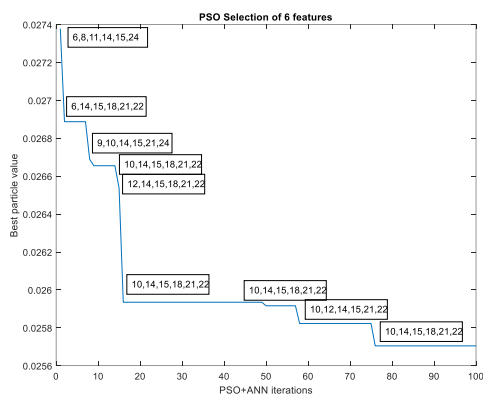
b) Tok selekcije 3 varijable



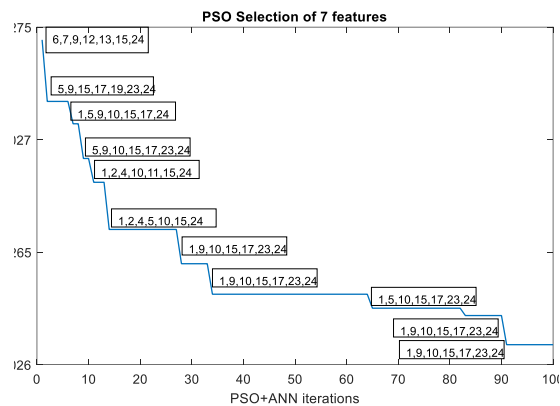
c) Tok selekcije 4 varijable



d) Tok selekcije 5 varijabli



e) Tok selekcije 6 varijabli



f) Tok selekcije 7 varijabli

Slika 5.11 Ilustracija konvergencije PSO+ANN algoritma prilikom selekcije obeležja

Analiza rezultata PSO+ANN algoritma selekcije obeležja je pokazala da je najbolji podskup od pet varijabli [Pol, Pojavni oblik ponašanja, Starosne grupe, Pesački\_prelaz, Prelaženje kolovoza] sa indeksima obeležja [9, 11, 20, 24, 25]. Ovih

### *Povećanje razumljivosti modela zasnovanog na pravilima*

pet varijabli će se koristiti kao ulazi generisanih modela zasnovanih na pravilima. Analiza rezultata selekcije obeležja zasnovana na kros-korelacionim matricama pokazala se korisnom za eliminaciju redundantnih varijabli.

## 6. Rezultati metodologije

Poglavlje sadrži detaljni opis dva TSK fazi modela i modela asocijativnih pravila koji su svi generisani za istu bazu podataka saobraćajnih nezgoda sa učešćem pešaka. Dati su rezultati razvijenog okvira za povećanje razumljivosti ovih modela pomoću detaljnih objašnjenja pronađenih relacija između TSK fazi pravila i asocijativnih pravila. Prikazani su i grafički izlazi primenjene metodologije koji takođe pomažu u razumevanju modela. Na kraju su upoređeni kvantitativni pokazatelji interpretabilnosti sa drugim sličnim modelima zasnovanim na pravilima.

### 6.1 Generisanje asocijativnih pravila Apriori algoritmom

Pomoću opisanog Apriori algoritma generisana su asocijativna pravila kojima se tumači veza između vrednosti raznih atributa koji se zajedno pojavljuju u uzorcima baze podataka koji opisuju određene događaje - saobraćajne nezgode. Uzimajući u obzir da je nivo pouzdanosti (eng. *confidence level*) postavljen da bude iznad 0.6, generisano je ukupno 45 asocijativnih pravila. Posledični deo pravila sadrži izlaznu varijablu *Posledice* koja označava težinu udesa. Najčešća vrednost varijable *Posledice* koja se pojavljuje u generisanom skupu asocijativnih pravila je sa značenjem „Bez povreda“. Vrednosti koje odgovaraju značenju „Sa većim povredama“ se pojavljuju u samo 14% slučajeva celokupnog broja uzoraka, pa se iz tog razloga ove vrednosti nisu pojavile u generisanom skupu asocijativnih pravila.

Generisana, a potom i selektovana asocijativna pravila su prikazana u Tabeli 6.1. Broj generisanih pravila je dodatno redukovano na osnovu dva kriterijuma: nivoa pouzdanosti i broja ulaznih atributa. Generisana pravila su prioritizovana na osnovu izračunatog nivoa pouzdanosti asocijativnog pravila koji je dobijen od Apriori algoritma. Potom je urađeno filtriranje asocijativnih pravila koja su međusobno slična po specifičnom kriterijumu koristeći algoritam povezivanja asocijativnih i TSK fazi pravila. Naime, algoritam povezivanja asocijativnih i TSK fazi pravila isključuje pravila koja predstavljaju podskup istih varijabli i sa istom vrednošću. Na taj način se dobija manji skup asocijativnih pravila sa specifičnijim uslovima, što olakšava razvoj algoritma i interpretaciju rezultata. Na primer, ako dva pravila imaju varijablu *Pol* sa značenjem „Muški“, a prvo pravilo sadrži još i varijablu *Starosne grupe* „15-29“, drugo jednostavnije pravilo neće biti razmatrano.

Tabela 6.1 Odabrana asocijativna pravila

Asoc. pravilo	Uzroci saobraćajne nesreće	Broj pojavljivanja	Posledice saobraćajne nesreće	Br.	Pouzdanost
1	<i>Pol</i> ="Muški" <i>Pešački prelaz</i> ="Na pešačkom prelazu" <i>Prelaženje kolovoza</i> ="Ostalo"	248	„Bez povreda“	240	0.97
2	<i>Pol</i> ="Muški" <i>Prelaženje kolovoza</i> ="Ostalo" <i>Starosne grupe</i> ="45-64"	182	„Bez povreda“	166	0.91
3	<i>Pol</i> ="Muški" <i>Prelaženje kolovoza</i> ="Ostalo" <i>Starosne grupe</i> ="30-44"	212	„Bez povreda“	189	0.89
4	<i>Pešački prelaz</i> ="Van pešačkog prelaza" <i>Prelaženje kolovoza</i> ="Ostalo"	397	„Bez povreda“	298	0.75
5	<i>Pol</i> ="Ženski" <i>Pešački prelaz</i> ="Na pešačkom prelazu" <i>Ponašanje</i> ="Prolazak kroz crveno svetlo"	214	„Sa manjim povredama“	159	0.74
6	<i>Prelaženje kolovoza</i> ="Sa leve na desnu stranu"	271	„Sa manjim povredama“	191	0.70
7	<i>Ponašanje</i> ="Prolazak kroz crveno svetlo" <i>Starosne grupe</i> ="15-29"	289	„Sa manjim povredama“	180	0.62

Odabrana asocijativna pravila, uključujući promenljive i vrednosti promenljivih koje u njima figurišu su data u Tabeli 6.2. Sve su varijable definisane u jedinstvenom intervalu razmatranja u granicama [0,1]. Interval razmatranja je podeljen u L-1 sekcija, gde je L broj vrednosti date varijable. Dakle, tabela sadrži kolone sa nazivima varijabli, sa njihovim lingvističkim i numeričkim vrednostima, dok preostale kolone označavaju koja asocijativna pravila sadrže varijablu sa datom vrednošću [102]. Na primer, varijabla *Pol* sa vrednošću "žensko" se pojavljuje u asocijativnom pravilu broj 5.



Tabela 6.2 Definicija funkcija pripadnosti normalizovanog prostora razmatranja varijabli asocijativnih pravila

Varijabla	Značenje	Normalizovana vrednost	Asocijativno pravilo						
			1	2	3	4	5	6	7
Ponašanje	„Prolazak kroz crveno svetlo“	0.0					x		x
	„Pretrčavanje kolovoza“	0.167							
	„Nepravilno kretanje unazad“	0.333							
	„Neodgovarajuća brzina“	0.5							
	„Nepoštovanje prava prvenstva“	0.667							
	„Kretanje u suprotnom pravcu“	0.833							
	„Vožnja pod uticajem narkotika“	1.0							
Pol	„Nepoznat“	0.0							
	„Muški“	0.5	x	x	x				
	„Ženski“	1.0					x		
Starosne grupe	„Nepoznata“	0.0							
	„0-14“	0.2							
	„15-29“	0.4							x
	„30-44“	0.6			x				
	„45-64“	0.8		x					
	„65>“	1.0							
Pešački prelaz	„Na pešačkom prelazu“	0.0	x				x		
	„Van pešačkog prelaza“	0.33			x				
	„Trotoar“	0.66							
	„Ostalo“	1.0							
Prelaženje kolovoza	„Sa leve na desnu stranu“	0.0						x	
	„Sa desne na levu stranu“	0.5							
	„Ostalo“	1.0	x		x	x			
Posledice	„Poginuli“	0.0							
	„Sa većim povredama“	0.33							
	„Sa manjim povredama“	0.66					x	x	x
	„Bez povreda“	1.0	x	x	x	x			

U tekstu ispod su detaljno opisana dva generisana TSK fazi modela. Izborom dva TSK fazi modela želelo se proceniti kako broj pravila i broj parametara TSK modela utiče na njihovu interpretabilnost. Takođe, razmotriće se i važnost pojedinih asocijativnih pravila i TSK fazi pravila, kao i njihov međusobni uticaj.

Broj parametara TSK fazi modela se računa kao  $2K(N+1)$ , gde je  $K$  broj pravila (klastera), a  $N$  je broj ulaza [93]. Prvi TSK fazi model sadrži pet ulaza ( $N=5$ ), dok drugi TSK fazi model poseduje tri ulaza ( $N=3$ ). Kako bi oba modela imala približno isti broj parametara, za TSK fazi model sa pet ulaza (TSK model 1) su generisana tri pravila, a za TSK fazi model sa tri ulaza (TSK model 2) je generisano pet pravila. Na osnovu toga, prvi TSK fazi model ima 36 parametara, a drugi TSK fazi model ima 40 parametara (približno isti broj parametara). Sa oba TSK fazi modela su postignute približno iste srednje kvadratne greške (eng. *Mean Square Error*), odnosno  $MSE_1=0.0327$  i  $MSE_2=0.0295$ , kao što je i očekivano zbog sličnog broja parametara. TSK fazi modeli su definisani sa  $K$  pravila datih u jednačini (3.18), kao i istim brojem odgovarajućih centara klastera datih u jednačini (3.22) i (3.24).

## 6.2 Prvi TSK fazi model

Prvi TSK fazi model poseduje pet ulaza i sadrži tri generisana pravila. Ukupno, prvi TSK model ima 36 parametara. U Tabeli 6.3 su prikazani parametri posledičnog dela prvog TSK modela, dok su u Tabeli 6.4 dati uslovni, odnosno uzročni delovi pravila, što uključuje koordinate centara klastera. Nagib ( $a$ ) i presek ( $b$ ) koeficijentata linearne jednačine čine posledični deo TSK modela.

Tabela 6.3 Posledični deo parametara

Pravilo	Ponašanje ( $a_1$ )	Pol ( $a_2$ )	Starosne grupe ( $a_3$ )	Pešački prelaz ( $a_4$ )	Prelaženje kolovoza ( $a_5$ )	Koef. ( $b$ )
Pravilo 1 ( $y_1$ )	0.701	-0.370	-0.156	-0.346	0.350	0.928
Pravilo 2 ( $y_2$ )	0.060	-0.042	-0.094	-0.059	-0.010	0.543
Pravilo 3 ( $y_3$ )	0.080	-0.045	0.009	-0.049	1.336	-0.377

Tabela 6.4 Koordinate centroida pravila

Koordinate centra klastera (Centroid)	Ponašanje ( $c_1$ )	Pol ( $c_2$ )	Starosne grupe ( $c_3$ )	Pešački prelaz ( $c_4$ )	Prelaženje kolovoza ( $c_5$ )	Posledice-težina udesa ( $y$ )
Klaster 1	0.012	0.795	0.644	0.553	0.977	0.615
Klaster 2	0.043	0.800	0.576	0.151	0.271	0.435
Klaster 3	0.395	0.528	0.538	0.202	0.994	0.994

Dobijeni parametri generisanog TSK fazi modela prikazuju uticaj pravila i ulaza na konačne vrednosti posledica u datim uslovima. Parametri nagiba pokazuju relativan značaj ulaznih varijabli odgovarajućih pravila, dok položaj centroida opisuje opseg primene pravila za odgovarajuće ulazne vrednosti.

Iz Tabele 6.3 sa posledičnim delom parametara se vidi da su najdominantnije vrednosti nagiba za varijable *Ponašanje* kod prvog pravila i za varijablu *Prelaženje kolovoza* kod trećeg pravila. Kod drugog pravila su vrednosti izbalansirane. Prva dva pravila dominantno definišu povrede kao posledice, dok treće pravilo opisuje posledice „Bez povreda“.

U algoritmu za povezivanje asocijativnih i TSK fazi pravila opisanom u poglavlju 4.5 proverava se intezitet preklapanja posledičnih delova asocijativnih i TSK fazi pravila. Ukoliko je intezitet ispod određenog opsega, provera preklapanja uslovnih delova asocijativnih i TSK fazi pravila se neće izvršiti.

Vrednost kumulativnog preklapanja između asocijativnog pravila i TSK fazi pravila predstavlja meru njihove sličnosti. Vrednost kumulativnog preklapanja se računa kao zbir pojedinačnih ulaznih varijabli TSK fazi pravila i uzročnih varijabli asocijativnih pravila podeljen sa brojem uzročnih varijabli.

Tabela 6.5 sadrži vrednosti preklapanja uslovnih varijabli asocijativnih i TSK fazi pravila, kao i njihovih posledičnih delova. Podebljane vrednosti u Tabeli 6.5 označavaju visoko preklapljene varijable (iznad 0.8). Varijable *Prelaženje kolovoza* i *Pol* se kod prvog TSK fazi modela pokazale kao najčešće aktivirane varijable u procesu povezivanja asocijativnih i TSK fazi pravila. Varijable *Ponašanje* i *Starosne grupe* su korišćene u mnogo manjem broju. Varijabla *Pešački prelaz* se pokazala kao najređe aktivirana varijabla, tj. najmanje je korišćena prilikom povezivanja pravila.

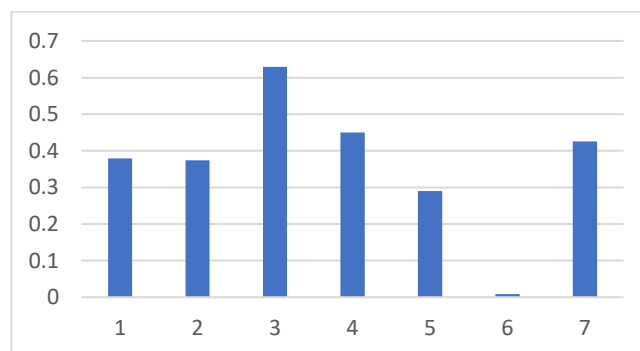
Tabela 6.5 Preklapanja asocijativnih pravila sa pravilima prvog TSK fazi modela

Asoc. pravilo	TSK pravilo	Uslovne varijable					Posledice – težina udesa
		<b>Ponašanje</b>	<b>Pol</b>	<b>Starosne grupe</b>	<b>Pešački prelaz</b>	<b>Prelaženje kolovoza</b>	
1	1		0.013			0.974	0.001
	2		0.011		0.32		
	3		<b>0.962</b>		0.13	<b>0.998</b>	<b>0.998</b>
2	1		0.013	0.296		0.974	0.001
	2		0.011	0.081			
	3		<b>0.962</b>	0.032		<b>0.998</b>	<b>0.998</b>
3	1		0.013	0.908		0.974	0.001
	2		0.011	0.972			

## Povećanje razumljivosti modela zasnovanog na pravilima

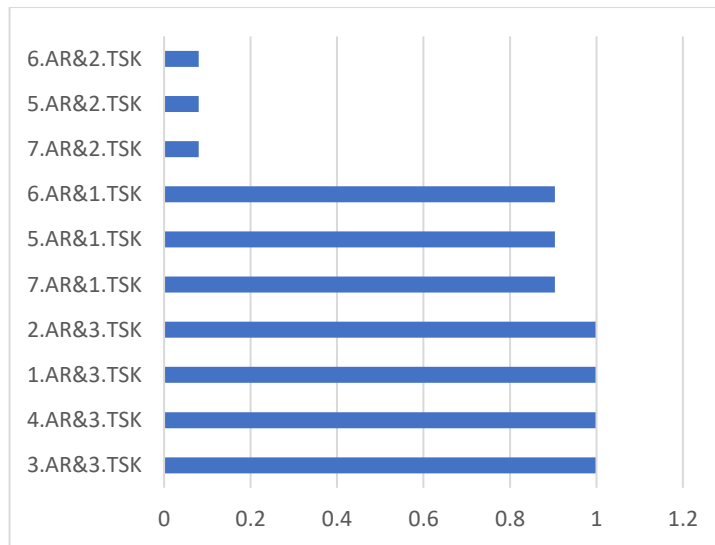
	3	<b>0.962</b>	<b>0.825</b>		<b>0.998</b>	<b>0.998</b>
	1			0.083	0.974	0.001
4	2			0.201		
	3			0.441	<b>0.998</b>	<b>0.998</b>
	1	0.993	0.122			<b>0.904</b>
5	2	0.912	0.135	0.32		0.08
	3			0.13		0.004
	1					<b>0.904</b>
6	2				0.025	0.08
	3					0.004
	1	<b>0.993</b>	0.051			<b>0.904</b>
7	2	0.912	0.213			0.08
	3		0.386			0.004

U cilju nalaženja asocijativnih pravila sa najsadržajnijim informacijama, iskorišćen je algoritam za interpretiranje asocijativnih pravila sa TSK fazi pravilima. U ovom slučaju je tražena sličnost asocijativnih pravila sa prvim TSK fazi modelom. Rezultat je prikazan na dijagramu na Slici 6.1. Dijagram na Slici 6.1 prikazuje nivo prisutnosti svakog pojedinačnog asocijativnog pravila u okviru prvog TSK fazi modela. Uočljivo je da su treće, četvrto i sedmo pravilo najzastupljenija u ovom TSK modelu, dok šesto asocijativno pravilo TSK model 1 ne sadrži.



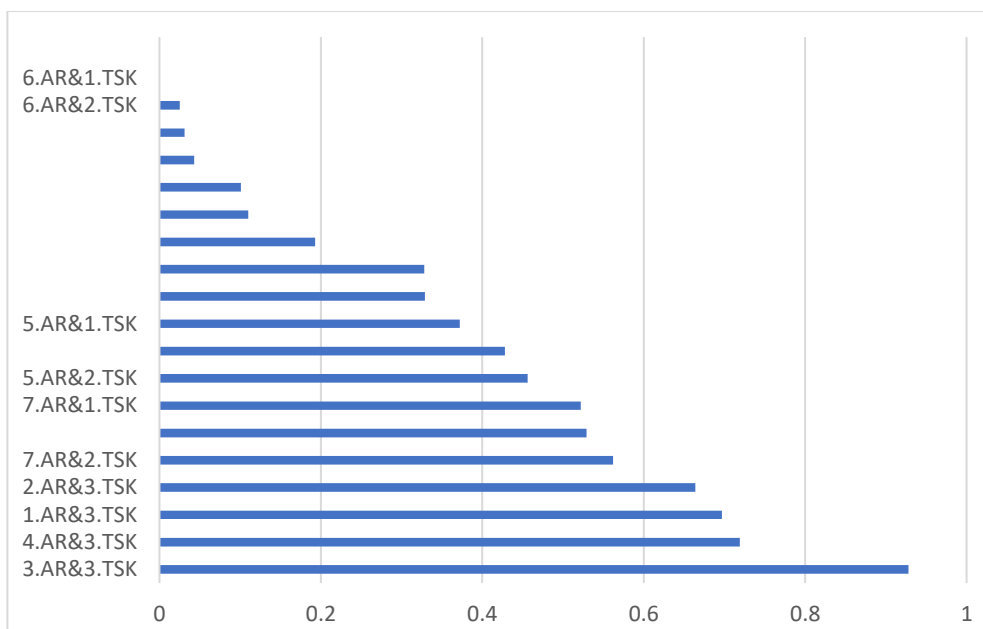
Slika 6.1 Kumulativna sličnost asocijativnih pravila i prvog TSK fazi modela

Slika 6.2 prikazuje intezitet preklapanja posledičnih delova asocijativnih i TSK fazi pravila prvog TSK fazi modela. U okviru algoritma za povezivanje asocijativnih i TSK pravila, gde se proverava nivo preklapanja posledičnih delova asocijativnih i TSK fazi pravila, odbacuju se parovi pravila koja ne ispunjavaju određeni prag. Na primer, za intezitet preklapanja koji je manji od 0.05, izvršavala bi se provera samo prvih 10 parova pravila od ukupno 21-og pravila.



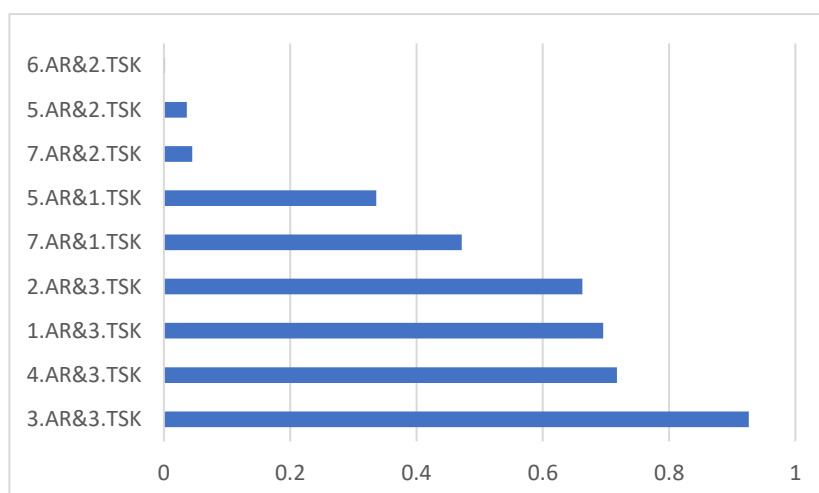
Slika 6.2 Intezitet preklapanja posledičnih delova parova asocijativnih i TSK pravila

Dijagram na Slici 6.3 opisuje nivo preklapanja uslovnih delova svakog asocijativnog pravila sa svim pravilima prvog TSK fazi modela. Stubići kod kojih nema natpisa para pravila odgovaraju parovima koji ne ispunjavaju uslov da je intezitet preklapanja posledičnih delova pravila iznad praga 0.05. Sa Slike 6.3 se vidi da je najviši nivo preklapanja između trećeg asocijativnog pravila i trećeg pravila prvog TSK fazi modela. Potom, po opadajućem intezitetu preklapanja, slede preklapanja trećeg pravila TSK fazi modela sa četvrtim, prvim i drugim asocijativnim pravilom. Sedmo asocijativno pravilo je preklapljeno sa prvim i drugim TSK pravilom, kao i peto asocijativno pravilo.



Slika 6.3 Preklapanja uslovnih delova svakog asocijativnog pravila sa svim pravilima prvog TSK fazi modela

Konačno, za parove asocijativnih i TSK pravila koje zadovoljavaju zadati prag preklapanja, izvršilo se množenje kumulativnih vrednosti preseka uslovnih delova asocijativnih i TSK fazi pravila sa odgovarajućim presecima posledičnih delova pravila. Sortirana lista proizvoda preseka pravila je prikazana na Slici 6.4. Redosled parova asocijativnih i TSK pravila sa najvećom sličnosti je ostala ista kao što je već navedeno u komentaru za prethodnu sliku.



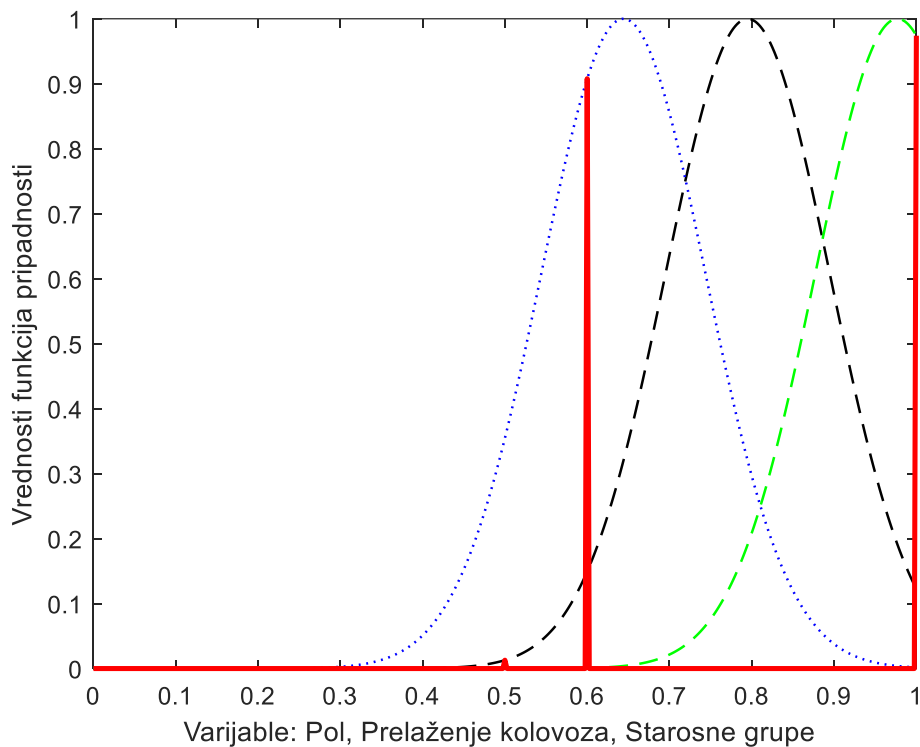
Slika 6.4 Sličnost asocijativnih pravila sa odgovarajućim pravilom prvog TSK fazi modela

Za detaljno objašnjavanje preklapanja asocijativnih pravila sa TSK fazi pravilima, postupak preklapanja će biti ilustriran na trećem, četvrtom, sedmom i prvom asocijativnom pravilu, jer su kod njih preklapanja najveća (videti Sliku 6.1).

Objašnjenja za TREĆE asocijativno pravilo koristeći TSK fazi model 1

Prostor razmatranja varijabli *Pol*, *Prelaženje kolovoza* i *Starosne grupe* je normalizovan u opsegu [0,1], što se vidi na Slikama 6.5, 6.6 i 6.7, a njihov položaj je kodovan po pravilima opisanim u tekstu uz Tabelu 6.2. U trećem asocijativnom pravilu varijabla *Pol* ima značenje „Muški“, što odgovara normalizovanoj vrednosti 0.5, varijabla *Prelaženje kolovoza* odgovara vrednosti „Ostalo“ (normalizovana vrednost 1), a varijabla *Starosne grupe* poseduje opseg „30-44“ (normalizovana vrednost 0.6).

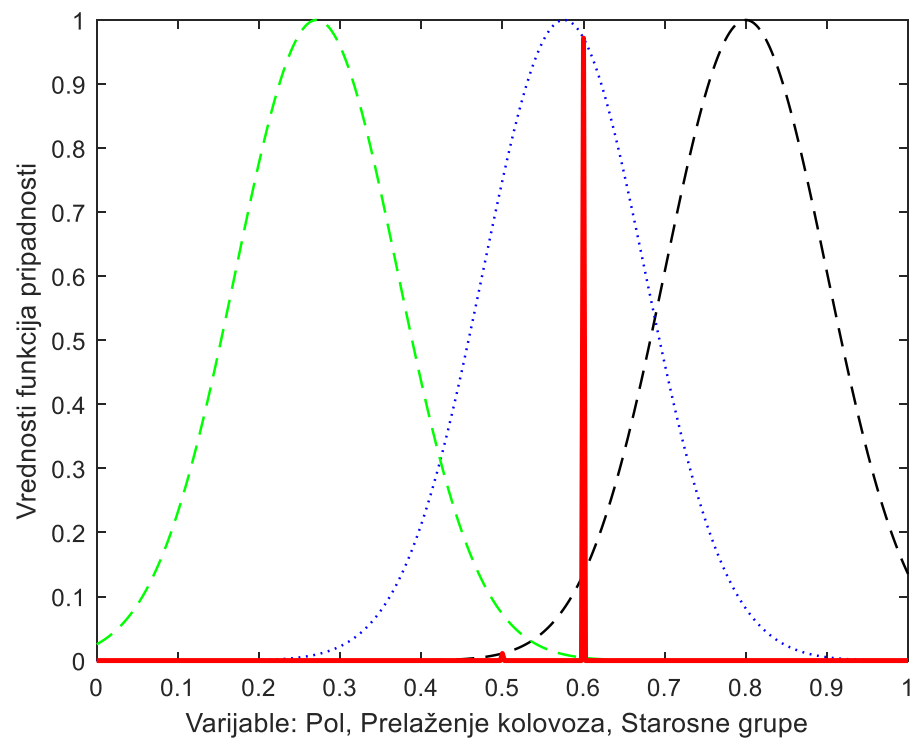
Kao što se vidi na Slici 6.5, za slučaj prvog TSK fazi pravila, funkcija pripadnosti za varijablu *Pol* ima centar u 0.795, funkcija pripadnosti varijable *Prelaženje kolovoza* ima centar u 0.977, a funkcija pripadnosti varijable *Starosne grupe* ima centar u 0.644. U slučaju varijable *Pol*, presek egzaktno vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila ima vrednost 0.013 (srednja Gausova funkcija prikazana isprekidanom linijom, presek u koordinati apscise 0), za varijablu *Prelaženje kolovoza* je presek egzaktno vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila (prikazane desnom isprekidanom linijom Gausove funkcije) u vrednosti 0.974, dok kod varijable *Starosne grupe* presek egzaktno vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila (prikazane tačkastom linijom Gausove funkcije) ima vrednost 0.908. Preseci funkcija pripadnosti za sve tri varijable su označeni na Slici 6.5 crvenim stubićima.



Slika 6.5 Presek uslovnih varijabli Pol, Prelaženje kolovoza i Starosne grupe

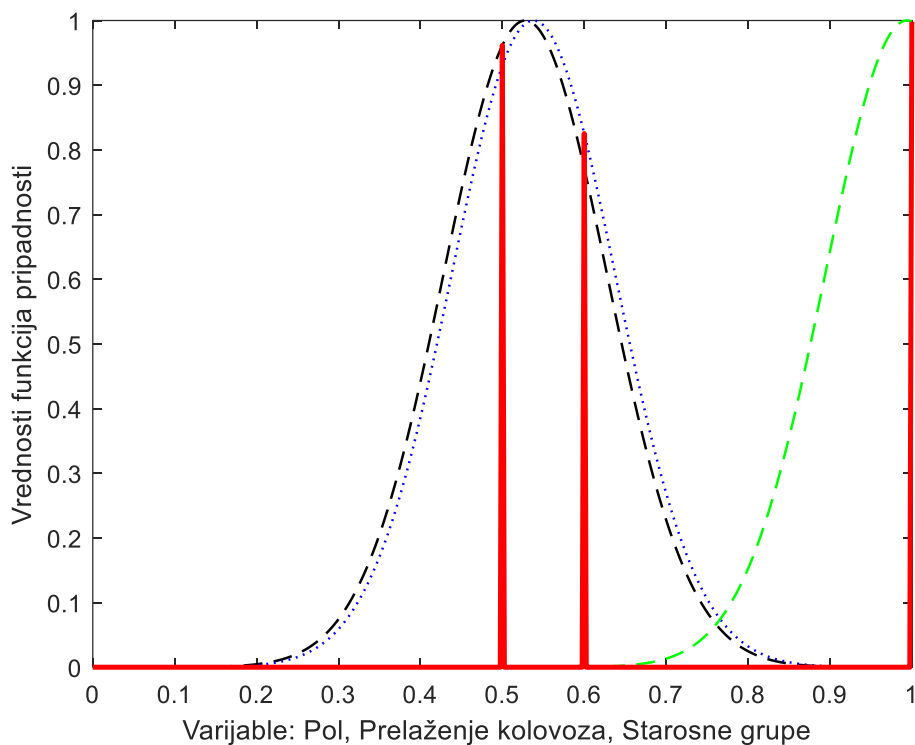
Kao što se vidi na Slici 6.6, za slučaj drugog TSK fazi pravila, funkcija pripadnosti za varijablu *Pol* ima centar u 0.8, funkcija pripadnosti varijable *Prelaženje kolovoza* ima centar u 0.271, dok funkcija pripadnosti varijable *Starosne grupe* ima centar u 0.576. U slučaju varijable *Pol*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK pravila ima vrednost 0.011 (u koordinati apscise 0), za varijablu *Prelaženje kolovoza* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK fazi pravila u vrednosti 0, dok kod varijable *Starosne grupe* presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK fazi pravila ima vrednost 0.972. Preseci funkcija pripadnosti za sve tri varijable su označeni na Slici 6.6 crvenim stubićima.





Slika 6.6 Presek uslovnih varijabli Pol, Prelaženje kolovoza i Starosne grupe

Kao što se vidi na Slici 6.7, za slučaj trećeg TSK fazi pravila, funkcija pripadnosti za varijablu *Pol* ima centar u 0.528, funkcija pripadnosti varijable *Prelaženje kolovoza* ima centar u 0.994, a funkcija pripadnosti za varijablu *Starosne grupe* ima centar u 0.538. U slučaju varijable *Pol*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti trećeg TSK fazi pravila ima vrednost 0.962 (u koordinati apscise 0), kod varijable *Prelaženje kolovoza* presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti trećeg TSK fazi pravila ima vrednost 0.998, dok je za varijablu *Starosne grupe* presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK fazi pravila u vrednosti 0.825. Preseci funkcija pripadnosti za sve tri varijable su označeni na Slici 6.7 crvenim stubićima.

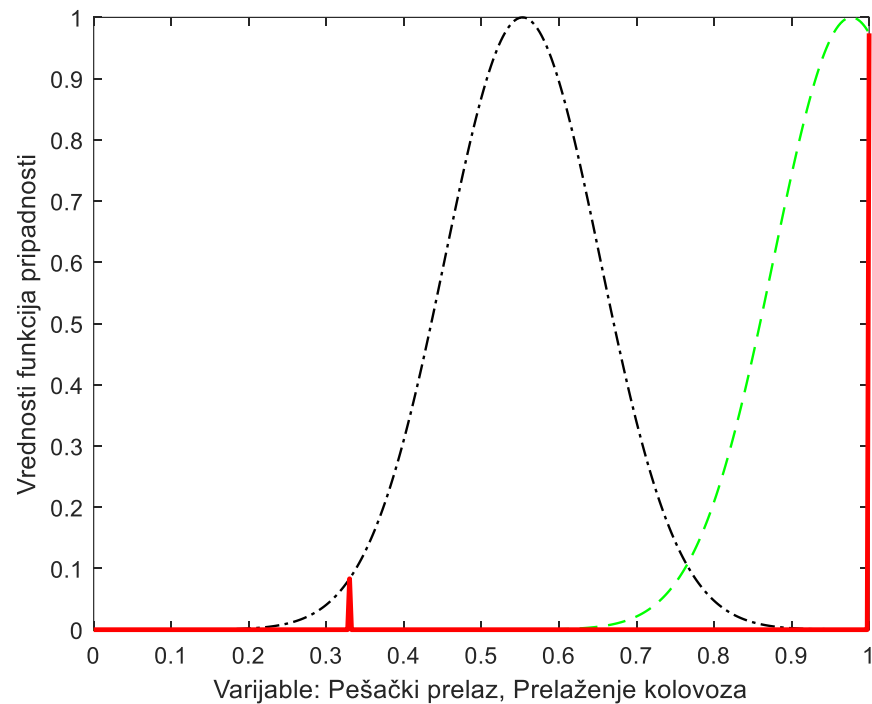


Slika 6.7 Presek uslovnih varijabli Pol, Prelaženje kolovoza i Starosne grupe

#### Objašnjenja za ČETVRTO asocijativno pravilo koristeći TSK fazi model 1

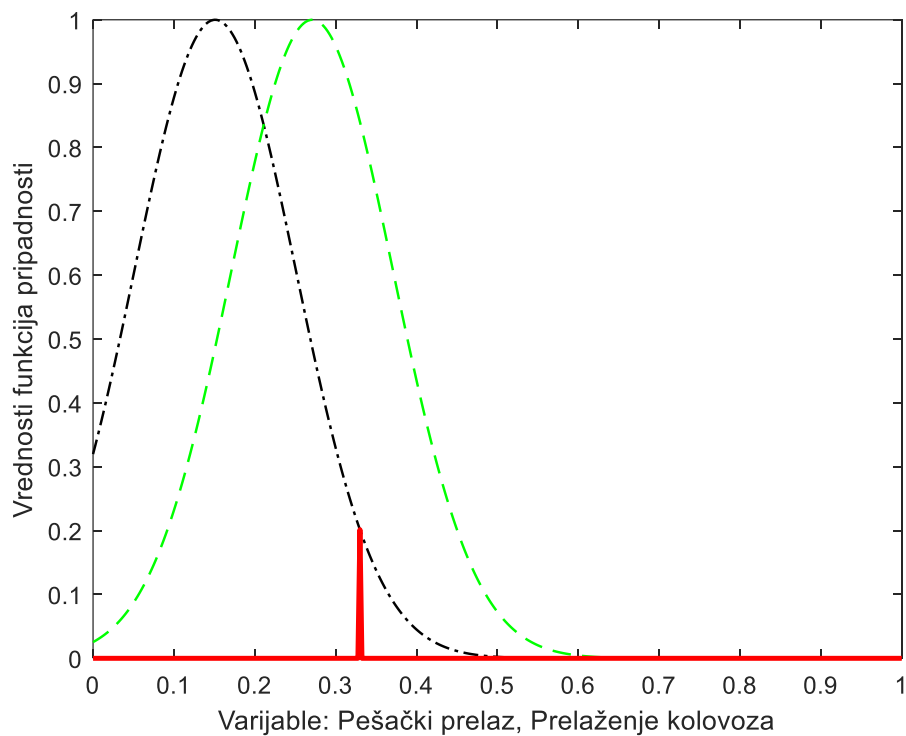
Prostor razmatranja varijabli *Pešački prelaz* i *Prelaženje kolovoza* je normalizovan u opsegu  $[0,1]$ , što se vidi na Slikama 6.8, 6.9 i 6.10, a njihov položaj je kodovan po pravilima opisanim u tekstu uz Tabelu 6.2. U četvrtom asocijativnom pravilu varijabla *Pešački prelaz* poseduje značenje „Van pešačkog prelaza“, što odgovara normalizovanoj vrednosti 0.33, a varijabla *Prelaženje kolovoza* odgovara vrednosti „Ostalo“ (normalizovana vrednost 1).

Kao što se vidi na Slici 6.8, za slučaj prvog TSK fazi pravila, funkcija pripadnosti za varijablu *Pešački prelaz* ima centar u 0.553, funkcija pripadnosti varijable *Prelaženje kolovoza* ima centar u 0.977. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila ima vrednost 0.083 (u koordinati apscise 0), a za varijablu *Prelaženje kolovoza* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila u vrednosti 0.974. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.8 crvenim stubićima.



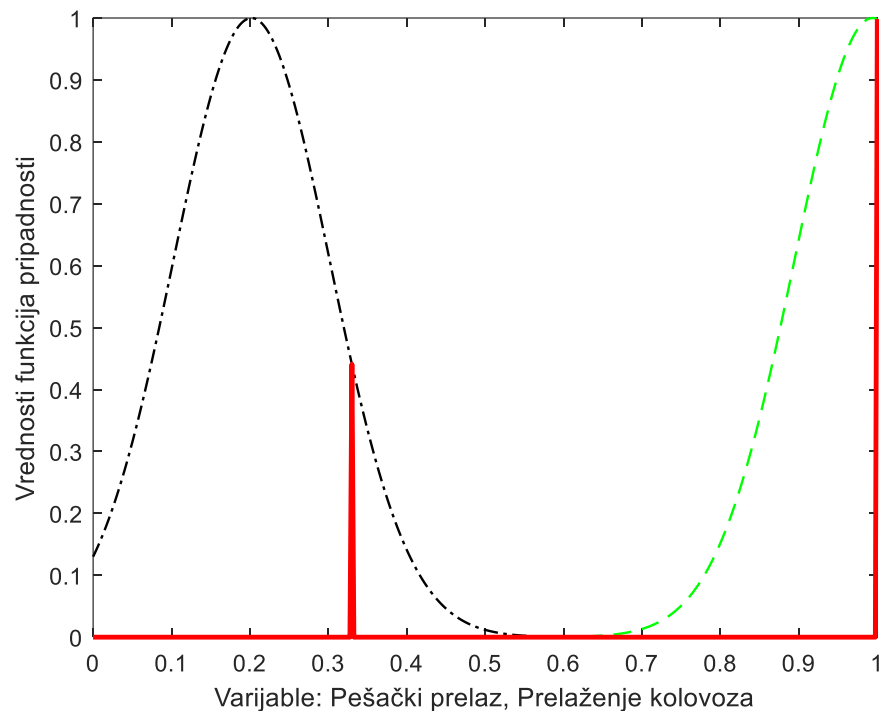
Slika 6.8 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza

Kao što se vidi na Slici 6.9, za slučaj drugog TSK fazi pravila, funkcija pripadnosti za varijablu *Pešački prelaz* ima centar u 0.151, a funkcija pripadnosti za varijablu *Prelaženje kolovoza* ima centar u 0.271. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila ima vrednost 0.201 (u koordinati apscise 0), a za varijablu *Prelaženje kolovoza* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila u vrednosti 0. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.9 crvenim stubićem.



Slika 6.9 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza

Kao što se vidi na Slici 6.10, za slučaj trećeg TSK fazi pravila, funkcija pripadnosti za varijablu *Pešački prelaz* ima centar u 0.202, a funkcija pripadnosti za varijablu *Prelaženje kolovoza* ima centar u 0.994. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila ima vrednost 0.441 (u koordinati apscise 0), a za varijablu *Prelaženje kolovoza* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila u vrednosti 0.998. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.10 crvenim stubićima.

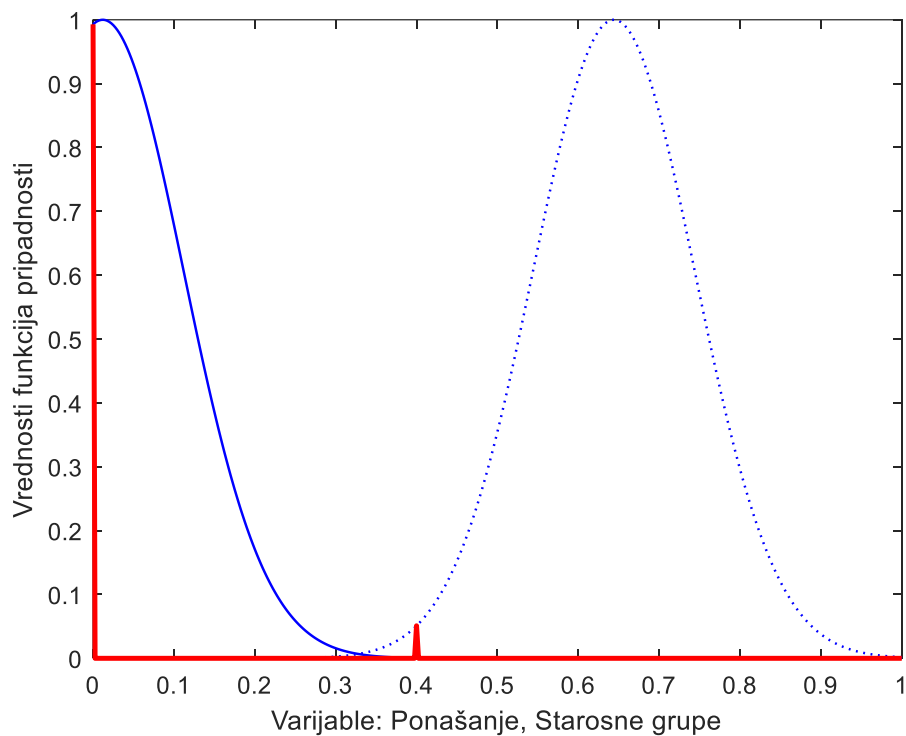


Slika 6.10 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza

#### Objašnjenja za SEDMO asocijativno pravilo pomoću TSK fazi modela 1

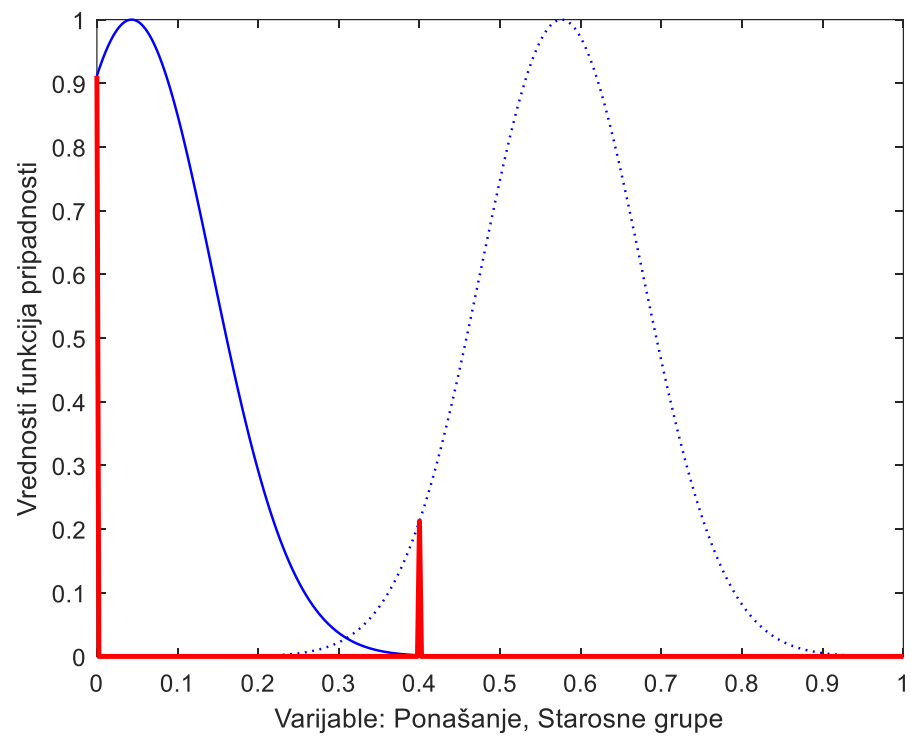
Prostor razmatranja varijabli *Ponašanje* i *Starosne grupe* je normalizovan u opsegu [0,1], što se vidi na Slikama 6.11, 6.12 i 6.13, a njihov položaj je kodovan po pravilima opisanim u tekstu uz Tabelu 6.2. U sedmom asocijativnom pravilu varijabla *Ponašanje* ima značenje „Prolazak kroz crveno svetlo“, što odgovara normalizovanoj vrednosti 0, dok varijabla *Starosne grupe* odgovara opsegu godina „15-29“ (normalizovana vrednost 0.4).

Kao što se vidi na Slici 6.11, za slučaj prvog TSK fazi pravila, funkcija pripadnosti za varijablu *Ponašanje* ima centar u 0.012, a funkcija pripadnosti varijable *Starosne grupe* ima centar u 0.644. U slučaju varijable *Ponašanje*, presek egzaktno vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila ima vrednost 0.993 (u koordinati apscise 0), dok kod varijable *Starosne grupe* presek egzaktno vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila ima vrednost 0.051. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.11 crvenim stubićima.



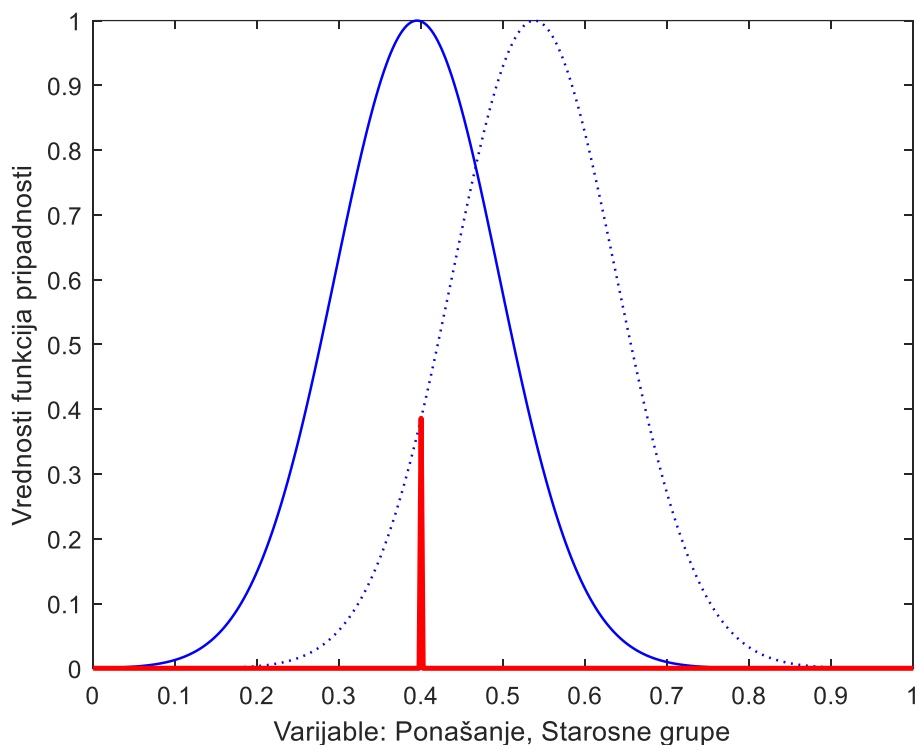
Slika 6.11 Presek uslovnih varijabli Ponašanje i Starosne grupe

Kao što se vidi na Slici 6.12, za slučaj drugog TSK fazi pravila, funkcija pripadnosti za varijablu *Ponašanje* ima centar u 0.043, a funkcija pripadnosti varijable *Starosne grupe* ima centar u 0.576. U slučaju varijable *Ponašanje*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK fazi pravila ima vrednost 0.912 (u koordinati apscise 0), dok kod varijable *Starosne grupe* presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK fazi pravila ima vrednost 0.213. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.12 crvenim stubićima.



Slika 6.12 Presek uslovnih varijabli Ponašanje i Starosne grupe

Kao što se vidi na Slici 6.13, za slučaj trećeg TSK fazi pravila, funkcija pripadnosti za varijablu *Ponašanje* ima centar u 0.395, a funkcija pripadnosti varijable *Starosne grupe* ima centar u 0.358, što se vidi na Slici 6.13. U slučaju varijable *Ponašanje* presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti trećeg TSK fazi pravila ima vrednost 0 (u koordinati apscise 0), dok varijabla *Starosne grupe* presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti trećeg TSK fazi pravila ima vrednost 0.386, što je označeno na Slici 6.13 crvenim stubićem.



Slika 6.13 Presek uslovnih varijabli Ponašanje i Starosne grupe

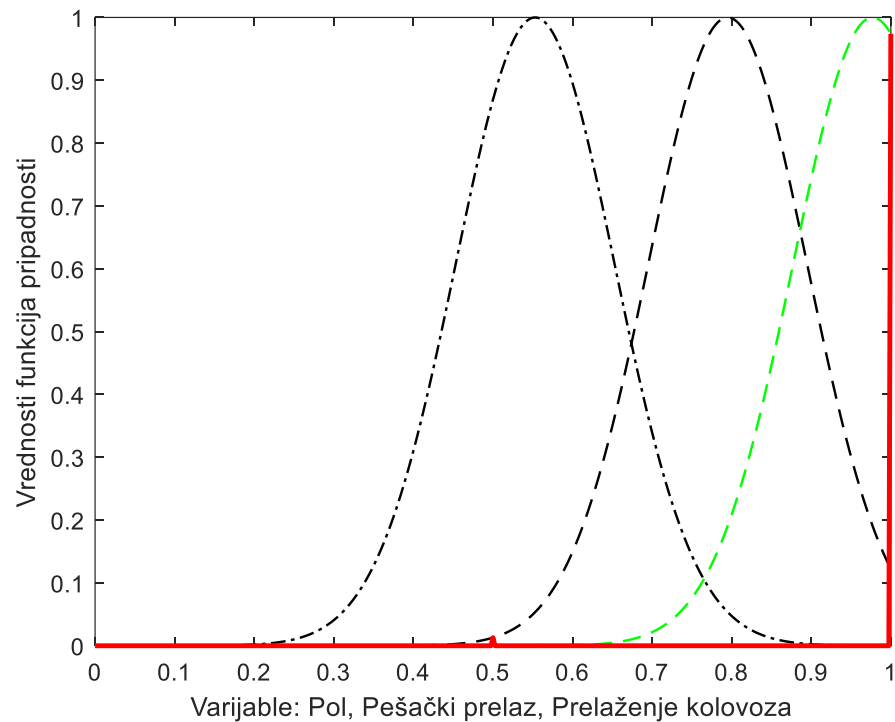
Objašnjenja za PRVO asocijativno pravilo koristeći TSK fazi model 1

Prostor razmatranja varijabli *Pol*, *Pešački prelaz* i *Prelaženje kolovoza* je normalizovan u opsegu [0,1], što se vidi na Slikama 6.14, 6.15 i 6.16, a njihov položaj je kodovan po pravilima opisanim u tekstu uz Tabelu 6.2. U prvom asocijativnom pravilu varijabla *Pol* ima značenje „Muški“, što odgovara normalizovanoj vrednosti 0.5, varijabla *Pešački prelaz* poseduje značenje „Na pešačkom prelazu“ (normalizovana vrednost 0), dok varijabla *Prelaženje kolovoza* odgovara vrednosti „Ostalo“ (normalizovana vrednost 1).

Kao što se vidi na Slici 6.14, za slučaj prvog TSK fazi pravila, funkcija pripadnosti za varijablu *Pol* ima centar u 0.795, funkcija pripadnosti varijable *Pešački prelaz* ima centar u 0.553, a funkcija pripadnosti varijable *Prelaženje kolovoza* ima centar u 0.977. U slučaju varijable *Pol*, presek egzaktne vrednosti (eng. *Singleton*) asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila ima vrednost 0.013 (u koordinati apscise 0), za varijablu *Pešački prelaz* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila u vrednosti 0.0, dok kod varijable *Prelaženje kolovoza* presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK

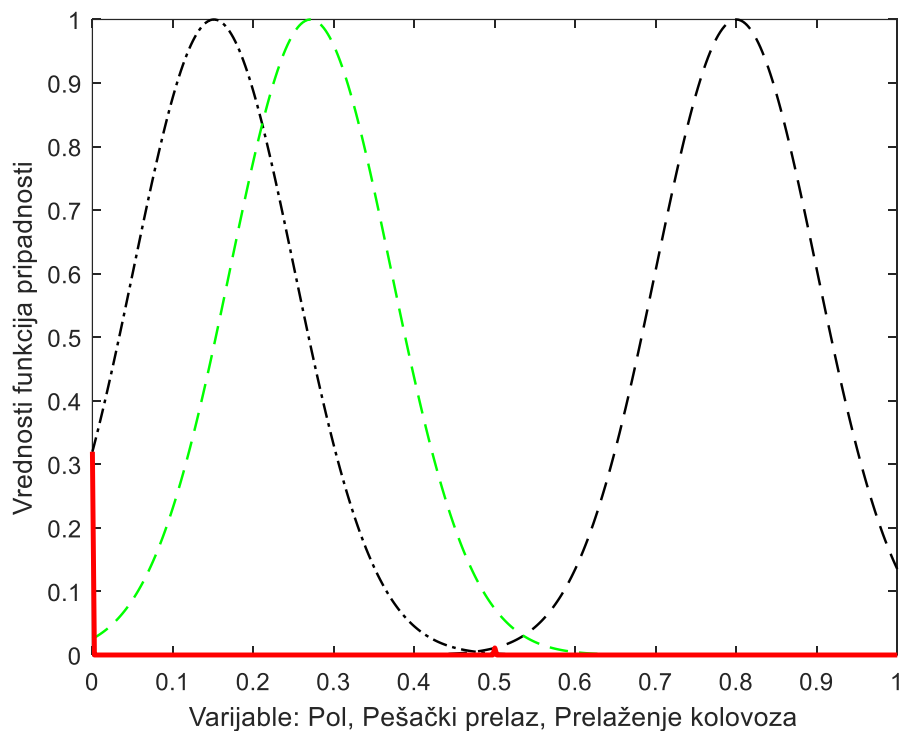


pravila ima vrednost 0.974. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.14 crvenim stubićima.



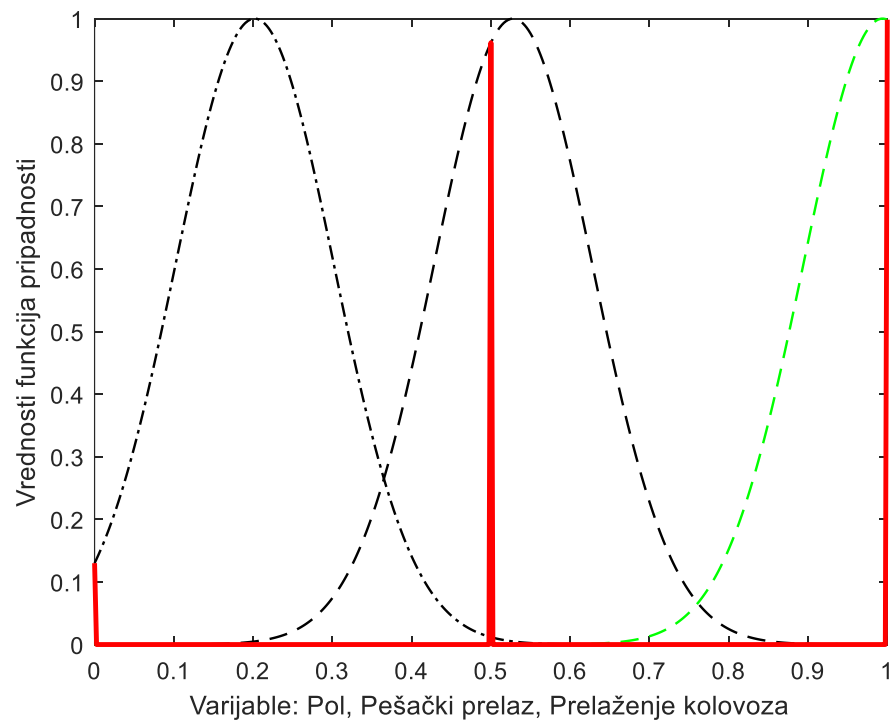
Slika 6.14 Presek uslovnih varijabli Pol, Pešački prelaz i Prelaženje kolovoza

Kao što se vidi na Slici 6.15, za slučaj drugog TSK fazi pravila, funkcija pripadnosti za varijablu *Pol* ima centar u 0.8, funkcija pripadnosti varijable *Pešački prelaz* ima centar u 0.151, a funkcija pripadnosti varijable *Prelaženje kolovoza* ima centar u 0.271. U slučaju varijable *Pol*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK pravila ima vrednost 0.011 (u koordinati apscise 0), za varijablu *Pešački prelaz* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK fazi pravila u vrednosti 0.320, dok kod varijable *Prelaženje kolovoza* presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK fazi pravila ima vrednost 0. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.15 crvenim stubićima.



Slika 6.15 Presek uslovnih varijabli Pol, Pešački prelaz i Prelaženje kolovoza

Kao što se vidi na Slici 6.16, za slučaj trećeg TSK fazi pravila, funkcija pripadnosti za varijablu *Pol* ima centar u 0.528, funkcija pripadnosti varijable *Pešački prelaz* ima centar u 0.202, a funkcija pripadnosti varijable *Prelaženje kolovoza* ima centar u 0.994. U slučaju varijable *Pol*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti trećeg TSK pravila ima vrednost 0.962 (u koordinati apscise 0), za varijablu *Pešački prelaz* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK fazi pravila u vrednosti 0.130, dok kod varijable *Prelaženje kolovoza* presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti trećeg TSK fazi pravila ima vrednost 0.998. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.16 crvenim stubićima.

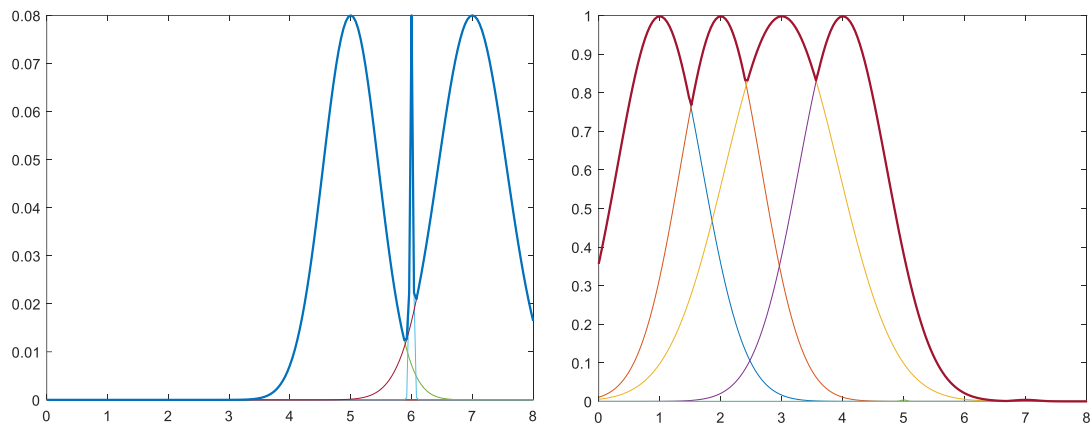


Slika 6.16 Presek uslovnih varijabli Pol, Pešački prelaz i Prelaženje kolovoza

Interesantno je analizirati preklapanje pojedinih TSK fazi pravila sa asocijativnim pravilima. Koristeći postupak preklapanja TSK modela sa asocijativnim pravilima objašnjen u poglavlju 4.5 dobijena je pokrivenost pravila prvog TSK fazi modela sa sedam selektovanih asocijativnih pravila. Postupak generisanja funkcija pokrivenosti fazi pravila sa asocijativnim pravilima kod prvog TSK fazi modela je ilustrovan na drugom i trećem TSK fazi pravilu, što je prikazano na Slici 6.17. Iz priloženih dijagrama je vidljiv način generisanja kumulativne funkcije pokrivenosti kombinovanjem više Gausovih funkcija preklapanja TSK fazi i asocijativnih pravila. Pokrivenost kompletnog prvog TSK fazi modela sa asocijativnim pravilima je prikazana na Slici 6.18. Za prvo TSK fazi pravilo je lako uočiti da se preklapa sa petim i sedmim asocijativnim pravilima. Veća širina funkcije kod sedmog asocijativnog pravila ilustruje veću pokrivenost ulaznog prostora nego kod petog asocijativnog pravila. Visina funkcije preklapanja pokazuje nivo preklapanja posledičnih delova pravila (TSK fazi pravila sa asocijativnim pravilima) i u ovom slučaju iznosi oko 0.8. Kod drugog TSK fazi pravila prisutna je pokrivenost sa petim, šestim i sedmim asocijativnim pravilom, međutim preklapanje posledica asocijativnih i TSK fazi pravila je vrlo nisko. Kod trećeg TSK fazi pravila je dobra pokrivenost sa prva četiri

## Povećanje razumljivosti modela zasnovanog na pravilima

asocijativna pravila i preklapljenost posledičnih delova pravila je maksimalna (vrednost 1.0).



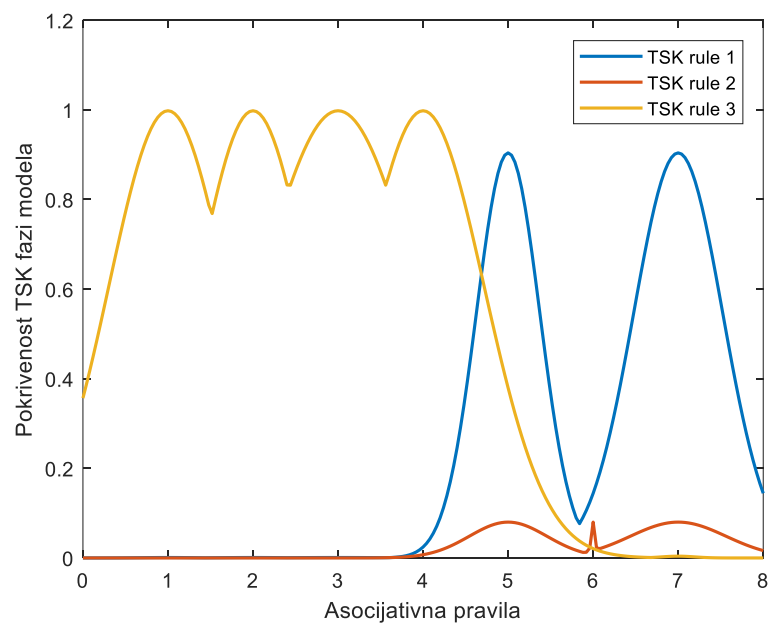
a) Generisanje funkcije pokrivenosti drugog fazi pravila prvog TSK fazi modela sa asocijativnim pravilom 5, 6

i 7

b) Generisanje funkcije pokrivenosti trećeg fazi pravila prvog TSK fazi modela sa prva četiri asocijativna pravila

pravila

Slika 6.17 Ilustracija postupka generisanja funkcija pokrivenosti TSK fazi pravila sa asocijativnim pravilima kod prvog TSK fazi modela



Slika 6.18 Pokrivenost pravila prvog TSK fazi modela sa asocijativnim pravilima

Zahvaljujući uspostavljenoj vezi između asocijativnih i pravila TSK fazi modela, moguće je interpretirati pravila prvog TSK fazi modela na sledeći način.

TSK fazi pravilo 1: Ovo pravilo predstavlja kombinaciju asocijativnih pravila 5 i 7. Posledica ovog pravila ima značenje „Sa manjim povredama“. U uslovnom delu pravila dominira varijabla *Ponašanje* sa vrednošću „Prolazak kroz crveno svetlo“. Dodatno, od uticaja su ženske osobe pri prelasku pešačkog prelaza i starosti od 15 do 29 godine.

TSK fazi pravilo 2: Pravilo predstavlja kombinaciju asocijativnih pravila 5, 6 i 7. Posledica ovog pravila takođe ima značenje „Sa manjim povredama“, ali sa mnogo manjim značajem. U uslovnom delu postoji i prelazak kolovoza sa leve na desnu stranu, uz prisustvo istih uticaja kao i u pravilu 1.

TSK fazi pravilo 3: Ovo pravilo je dominantno pravilo prvog TSK fazi modela i predstavlja kombinaciju prva četiri pravila. Posledica ovog pravila ima značenje „Bez povreda“. U uslovnom delu pravila dominira *Prelazak kolovoza* sa značenjem „Ostalo“, a potom osobe muškog pola. Pored toga, starost osoba u opsezima 30-44 i 45-64 godina i prelazak na ili izvan pešačkog prelaza.

Kada se uporede Slika 6.18 i Slika 6.4, može se primetiti potpuno slaganje rezultata dobijenih sa dva razmatrana algoritma–interpretera TSK fazi modela sa asocijativnim pravilima i interpretera asocijativnih pravila sa TSK fazi pravilom. Na primer, na Slici 6.4 se vidi da četiri najveće sličnosti postoje između trećeg TSK fazi pravila i prva četiri asocijativna pravila. Sa Slike 6.18 se takođe vidi ista ta zakonitost, odnosno da Gausova kriva koja opisuje pokrivenost trećeg TSK fazi pravila sa prva četiri asocijativna pravila pokriva najveću površinu i vrhovi Gausove krive su najvišeg inteziteta (blizu 1). Kriva koja opisuje TSK fazi pravilo 1 (na istoj Slici 6.18) i njenu zavisnost prema asocijativnim pravilima 7 i 5 ima nešto niže vrednosti vrhova krive. Shodno tome, na Slici 6.4 su naredna dva pravila po intezitetu sličnosti upravo pravila koja se odnose na prvo TSK fazi pravilo i asocijativna pravila 7 i 5. Pri tome je sedmo asocijativno pravilo značajnije od petog, što se takođe može potvrditi na Slici 6.18 jer je osnova Gausijana koja odgovara asocijativnom pravilu 7 šira nego kod pravila 5.

### 6.3 Drugi TSK fazi model

Drugi TSK fazi model poseduje tri ulaza za koje je generisano pet pravila. Pored toga, model poseduje ukupno 40 parametara. U Tabeli 6.6 su prikazani parametri posledičnog dela prvog TSK modela, dok su Tabelom 6.7 dati uslovni, odnosno uzročni delovi pravila, tj. koordinate centara klastera. Nagib ( $a$ ) i presek ( $b$ ) koeficijentata linearne jednačine čine posledični deo TSK fazi modela. Iz Tabele 6.6 sa posledičnim delom parametara se vidi da su najdominantnije vrednosti nagiba za varijable *Prelaženje kolovoza* kod prvog, drugog i petog pravila. Kod trećeg i četvrtog pravila su vrednosti uravnotežene. Prva dva pravila označavaju da su prvenstveno povezana sa posledicama „Bez posledica“, dok su treće i četvrto pravilo prvenstveno povezana sa saobraćajnim nesrećama sa povredama.

Tabela 6.6 Posledični deo parametara

Pravilo	Ponašanje ( $a_1$ )	Pešački prelaz ( $a_2$ )	Prelaženje kolovoza ( $a_3$ )	Koef. ( $b$ )
Pravilo 1 ( $y_1$ )	0.026	0.011	1.090	0.114
Pravilo 2 ( $y_2$ )	-0.050	-0.372	1.556	-0.611
Pravilo 3 ( $y_3$ )	0.100	-0.070	0.019	0.453
Pravilo 4 ( $y_4$ )	-0.428	0.135	-0.192	0.317
Pravilo 5 ( $y_5$ )	0.758	0.107	-1.945	2.497

Tabela 6.7 Koordinate centroida TSK fazi pravila

Koordinate centra klastera (Centroid)	Ponašanje ( $c_1$ )	Pešački prelaz ( $c_2$ )	Prelaženje kolovoza ( $c_3$ )	Posledice-težina udesa ( $y$ )
Klaster 1	0.698	0.155	0.994	0.992
Klaster 2	0.006	0.237	0.988	0.860
Klaster 3	0.038	0.132	0.276	0.432
Klaster 4	0.049	0.963	0.241	0.398
Klaster 5	0.079	0.819	1.000	0.681

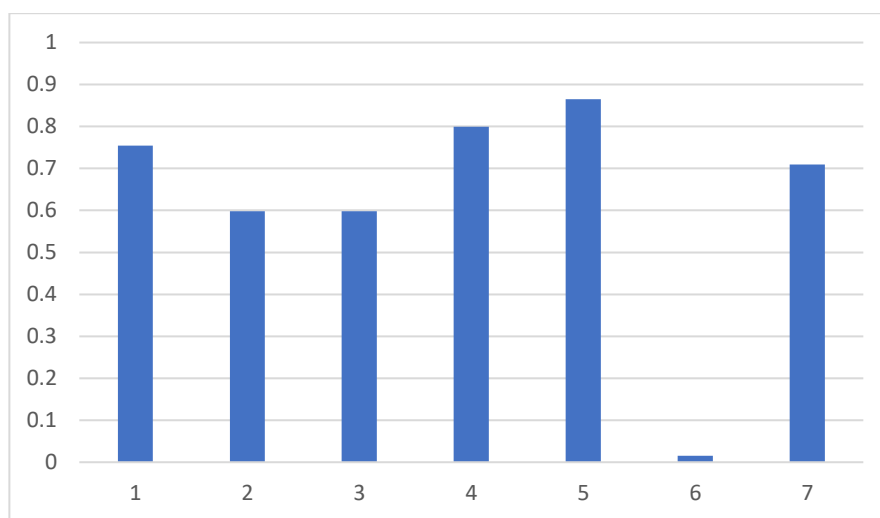
Tabela 6.8 sadrži vrednosti preklapanja uslovnih varijabli asocijativnih i TSK fazi pravila, kao i njihovih posledičnih delova. Podebljane vrednosti u Tabeli 6.8 označavaju visoko preklapljenе varijable (iznad 0.7). Varijabla *Prelaženje kolovoza* se kod drugog TSK fazi modela pokazala kao najčešće aktivirana varijabla u procesu

povezivanja asocijativnih i TSK fazi pravila. Varijabla *Pešački prelaz* se pokazala kao najređe aktivirana varijabla, tj. najmanje je korišćena prilikom povezivanja pravila.

Tabela 6.8 Preklapanja asocijativnih pravila sa pravilima drugog TSK fazi modela

Asoc. pravilo	TSK pravilo	Uslovne varijable			Posledice – težina udesa
		<i>Ponašanje</i>	<i>Pešački prelaz</i>	<i>Prelaženje kolovoza</i>	
<b>1</b>	1		0.301	<b>0.998</b>	<b>0.997</b>
	2		0.06	0.993	0.375
	3		0.418		
	4				
	5			1	0.006
<b>2</b>	1			<b>0.998</b>	<b>0.997</b>
	2			0.993	0.375
	3				
	4				
	5			1	0.006
<b>3</b>	1			<b>0.998</b>	<b>0.997</b>
	2			0.993	0.375
	3				
	4				
	5			1	0.006
<b>4</b>	1		0.216	<b>0.998</b>	<b>0.997</b>
	2		0.649	0.993	0.375
	3		0.141		
	4				
	5			1	0.006
<b>5</b>	1		0.301		0.004
	2	0.998	0.06		0.135
	3	0.93	0.418		0.074
	4	0.887			0.032
	5	<b>0.732</b>			<b>0.978</b>
<b>6</b>	1				0.004
	2				0.135
	3			0.022	0.074
	4			0.055	0.032
	5				<b>0.978</b>
<b>7</b>	1				0.004
	2	0.998			0.135
	3	0.93			0.074
	4	0.887			0.032
	5	<b>0.732</b>			<b>0.978</b>

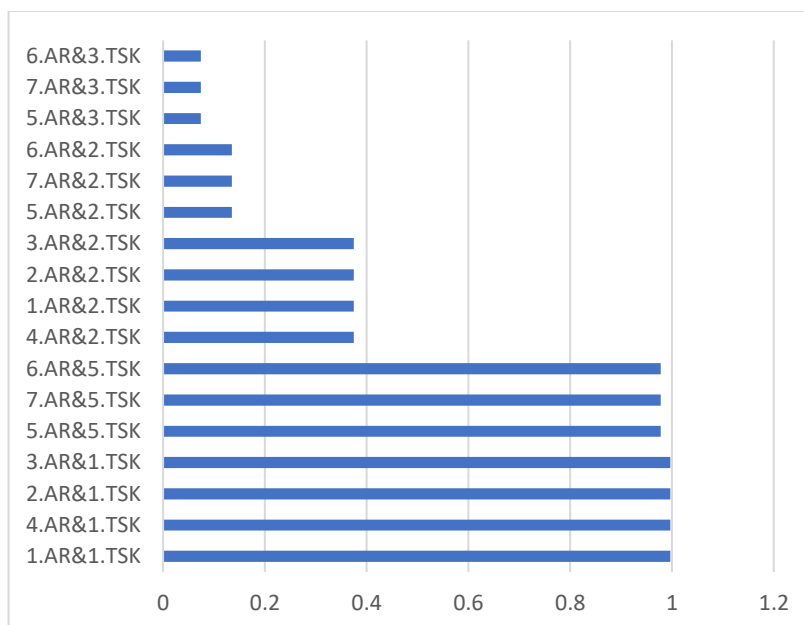
Dijagram na Slici 6.19 prikazuje nivo prisutnosti svakog pojedinačnog asocijativnog pravila u okviru drugog TSK fazi modela. Naime, pomenuta asocijativna pravila se identifikuju sabiranjem inteziteta preklapanja asocijativnih pravila sa TSK fazi pravilima i potom se dele sa brojem prisutnih uslovnih varijabli. Dakle, dobija se kumulativna sličnost asocijativnih pravila sa drugim TSK fazi modelom. Sa Slike 6.19 se vidi da su najdeskriptivnija, odnosno najbolje opisana pravila drugim TSK fazi modelom, pravila broj pet, četiri, jedan i sedam. Identično kao i kod prvog TSK fazi modela, šesto asocijativno pravilo nije zastupljeno ni u drugom TSK fazi modelu.



Slika 6.19 Kumulativna sličnost asocijativnih pravila i drugog TSK fazi modela

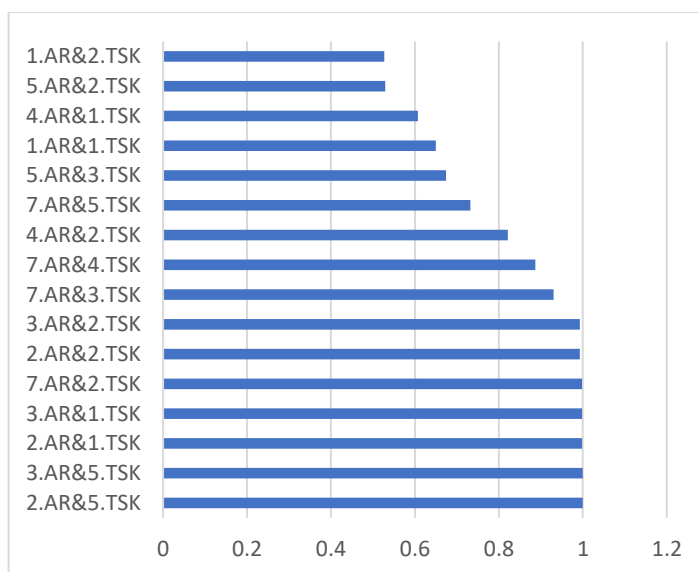
Slika 6.20 prikazuje intezitet preklapanja posledičnih delova asocijativnih i TSK fazi pravila. U algoritmu za povezivanje asocijativnih i TSK fazi pravila (Tabela 6.2) se proverava intezitet preklapanja posledičnih delova pravila. Ukoliko je intezitet ispod određenog opsega, provera preklapanja uslovnih delova asocijativnih i TSK fazi pravila se neće izvršiti. Na primer, za intezitet preklapanja koji je manji od 0.05, izvršavala bi se provera samo prvih 18 parova pravila od ukupno 35.





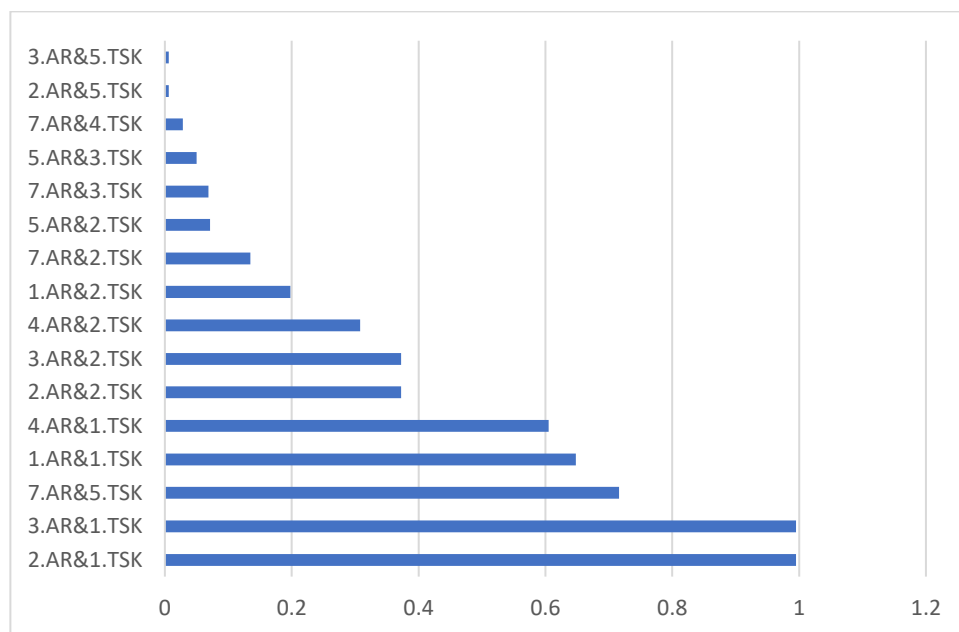
Slika 6.20 Intezitet preklapanja posledičnih delova asocijativnih i TSK fazi pravila

Dijagram na Slici 6.21 opisuje nivo preklapanja između svakog asocijativnog pravila i svih pravila drugog TSK fazi modela. Sa Slike 6.21 se vidi da je nivo preklapanja blizak jedinici čak kod sedam parova asocijativnih i TSK fazi pravila drugog TSK modela. Potom slede parovi pravila po ravnomerno opadajućem intezitetu preklapanja. Parovi pravila kod kojih nije ispunjen uslov praga preklapanja za posledične delove pravila nisu razmatrani.



Slika 6.21 Nivo preklapanja uslovnih delova asocijativnih pravila sa uslovnim delovima TSK fazi pravila

Konačno, za parove asocijativnih i TSK fazi pravila koji zadovoljavaju zadati prag preklapanja, izvršilo se množenje kumulativnih vrednosti preseka uslovnih delova asocijativnih i TSK fazi pravila sa odgovarajućim presecima posledičnih delova pravila. Sortirana lista proizvoda preseka pravila je prikazana na Slici 6.22. Dominantna sličnost je kod parova prvog TSK fazi pravila sa drugim i trećim asocijativnim pravilom. Vredi posebno naglasiti sličnost sedmog asocijativnog pravila sa petim TSK fazi pravilom, kao i sličnost prvog TSK fazi pravila sa prvim i četvrtim asocijativnim pravilom.



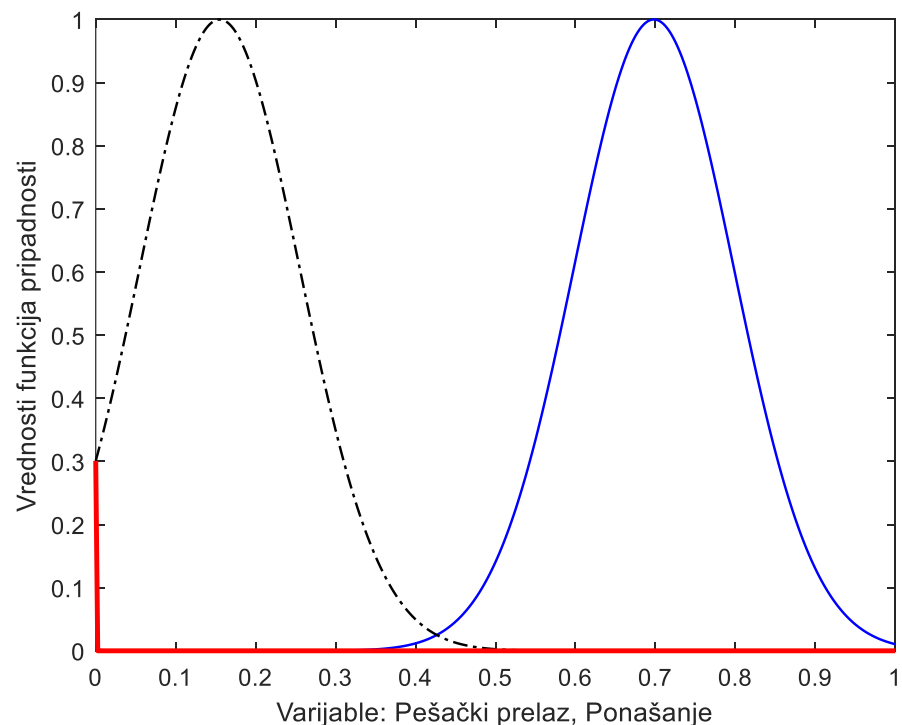
Slika 6.22 Sličnost asocijativnih pravila sa odgovarajućim pravilom drugog TSK fazi modela

U cilju objašnjenja preklapanja asocijativnih pravila sa TSK fazi pravilima, postupak preklapanja će biti ilustrovan asocijativnim pravilima broj pet i četiri sa obzirom da su kod njih preklapanja najveća (videti Sliku 6.19).

Objašnjenje za PETO asocijativno pravilo koristeći TSK fazi model 2

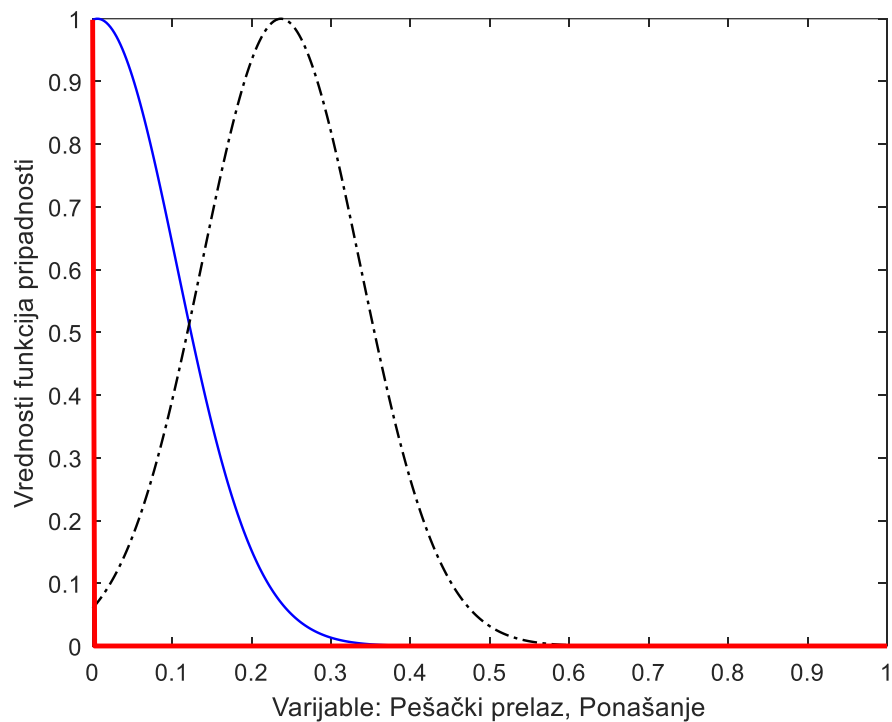
Prostor razmatranja varijabli *Pešački prelaz* i *Ponašanje* je normalizovan u opsegu [0,1], što se vidi na Slikama 6.23, 6.24, 6.25, 6.26 i 6.27, a njihov položaj je kodovan po pravilima opisanim u tekstu uz Tabelu 6.2. U petom asocijativnom pravilu varijabla *Pešački prelaz* poseduje značenje „Na pešačkom prelazu“ (normalizovana vrednost 0), dok varijabla *Ponašanje* odgovara vrednosti „Prolazak kroz crveno svetlo“ (normalizovana vrednost 0).

Kao što se vidi na Slici 6.23, za slučaj prvog TSK fazi pravila, funkcija pripadnosti varijable *Pešački prelaz* ima centar u 0.155, a funkcija pripadnosti varijable *Ponašanje* ima centar u 0.698. U slučaju varijable *Pešački prelaz* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila u vrednosti 0.301, dok kod varijable *Ponašanje* presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila ima vrednost nula. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.23 crvenim stubićem.



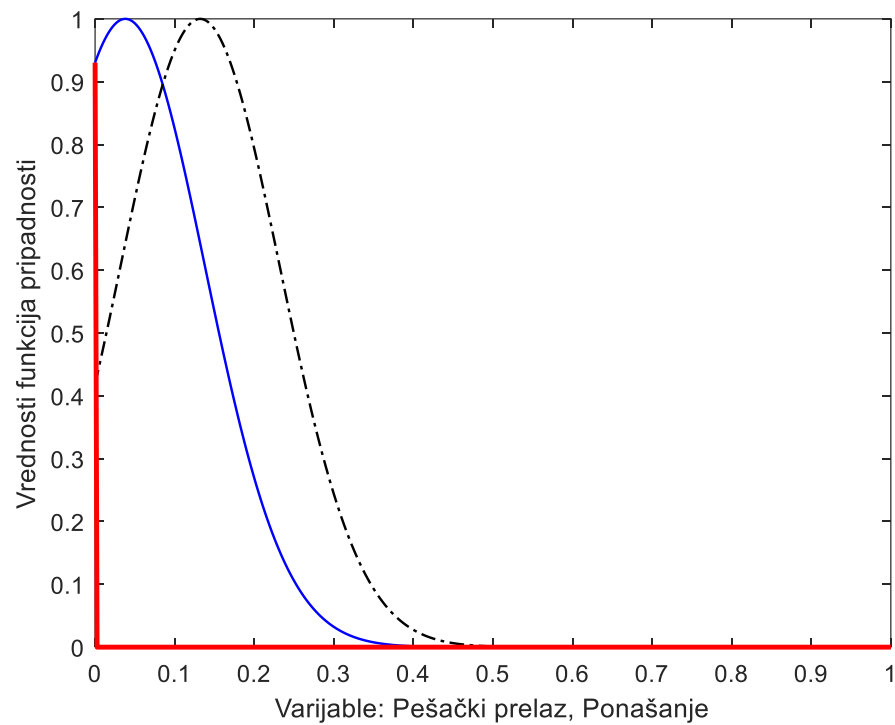
Slika 6.23 Presek uslovnih varijabli Pešački prelaz i Ponašanje

Kao što se vidi na Slici 6.24, za slučaj drugog TSK fazi pravila, funkcija pripadnosti za varijablu *Pešački prelaz* ima centar u 0.237, a funkcija pripadnosti za varijablu *Ponašanje* ima centar u 0.006. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK fazi pravila ima vrednost 0.060 (u koordinati apscise 0), a za varijablu *Ponašanje* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK fazi pravila u vrednosti 0.998. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.24 crvenim stubićem.



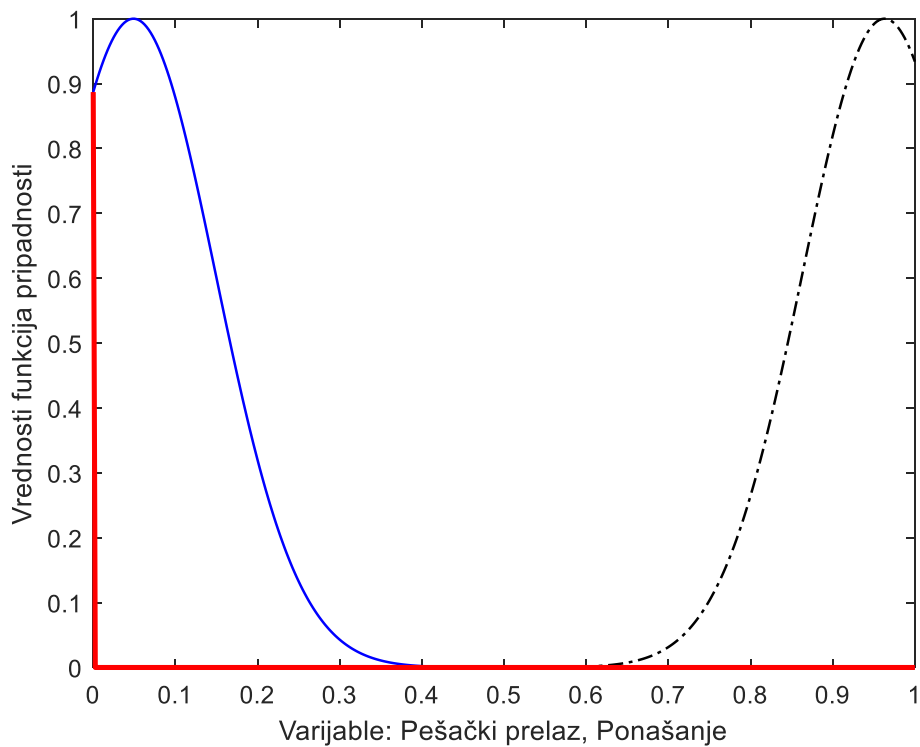
Slika 6.24 Presek uslovnih varijabli Pešački prelaz i Ponašanje

Kao što se vidi na Slici 6.25, za slučaj trećeg TSK fazi pravila, funkcija pripadnosti za varijablu *Pešački prelaz* ima centar u 0.132, a funkcija pripadnosti za varijablu *Ponašanje* ima centar u 0.038. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti trećeg TSK fazi pravila ima vrednost 0.418 (u koordinati apscise 0), a za varijablu *Ponašanje* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti trećeg TSK fazi pravila u vrednosti 0.93. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.25 crvenim stubićem.



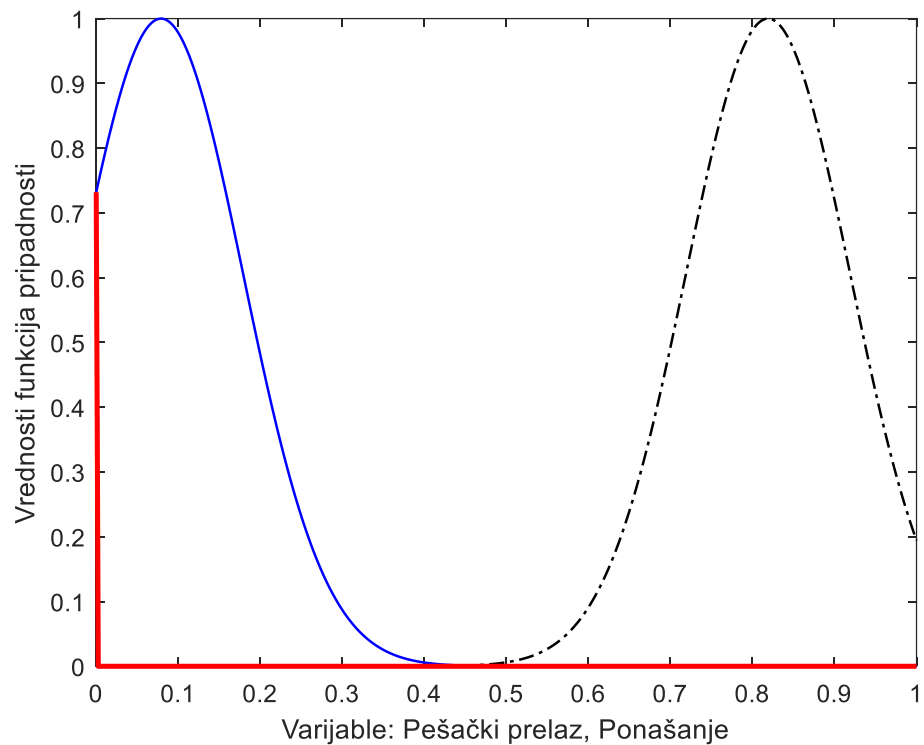
Slika 6.25 Presek uslovnih varijabli Pešački prelaz i Ponašanje

Kao što se vidi na Slici 6.26, za slučaj četvrtog TSK fazi pravila, funkcija pripadnosti za varijablu *Pešački prelaz* ima centar u 0.963, a funkcija pripadnosti za varijablu *Ponašanje* ima centar u 0.049. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti četvrtog TSK fazi pravila ima vrednost 0 (u koordinati apscise 0), a za varijablu *Ponašanje* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti četvrtog TSK fazi pravila u vrednosti 0.887. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.26 crvenim stubićem.



Slika 6.26 Presek uslovnih varijabli Pešački prelaz i Ponašanje

Kao što se vidi na Slici 6.27, za slučaj petog TSK fazi pravila, funkcija pripadnosti za varijablu *Pešački prelaz* ima centar u 0.819, a funkcija pripadnosti za varijablu *Ponašanje* ima centar u 0.079. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti petog TSK fazi pravila ima vrednost nula (u koordinati apscise 0), a za varijablu *Ponašanje* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti petog TSK fazi pravila u vrednosti 0.732. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.27 crvenim stubićem.

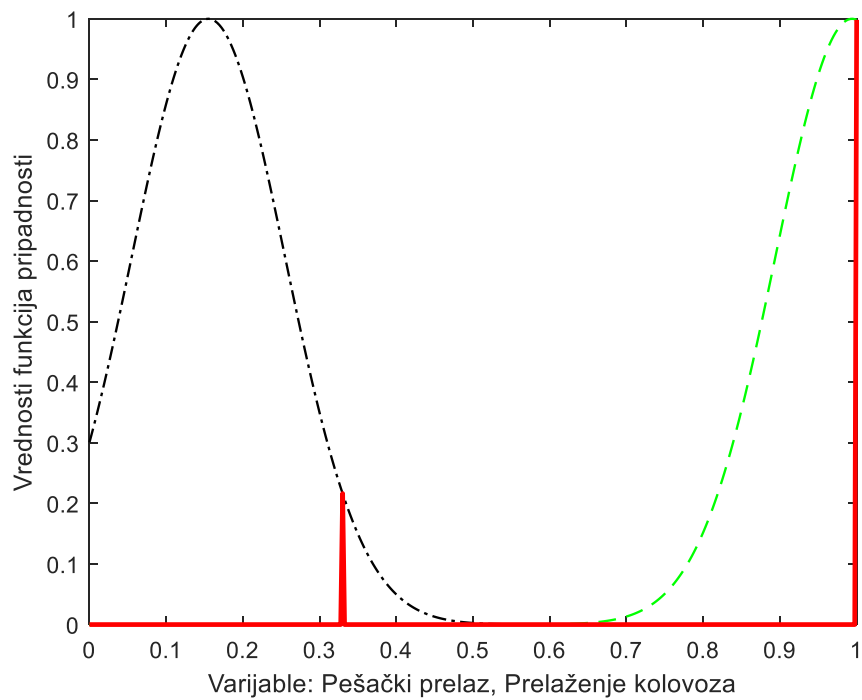


Slika 6.27 Presek uslovnih varijabli Pešački prelaz i Ponašanje

## Objašnjenje za ČETVRTO asocijativno pravilo koristeći TSK fazi model 2

Prostor razmatranja varijabli *Pešački prelaz* i *Prelaženje kolovoza* je normalizovan u opsegu  $[0,1]$ , što se vidi na Slikama 6.28, 6.29, 6.30, 6.31 i 6.32, a njihov položaj je kodovan po pravilima opisanim u tekstu uz Tabelu 6.2. U četvrtom asocijativnom pravilu varijabla *Pešački prelaz* poseduje značenje „Van pešačkog prelaza“ (normalizovana vrednost 0.33), dok varijabla *Prelaženje kolovoza* odgovara vrednosti „Ostalo“ (normalizovana vrednost 1).

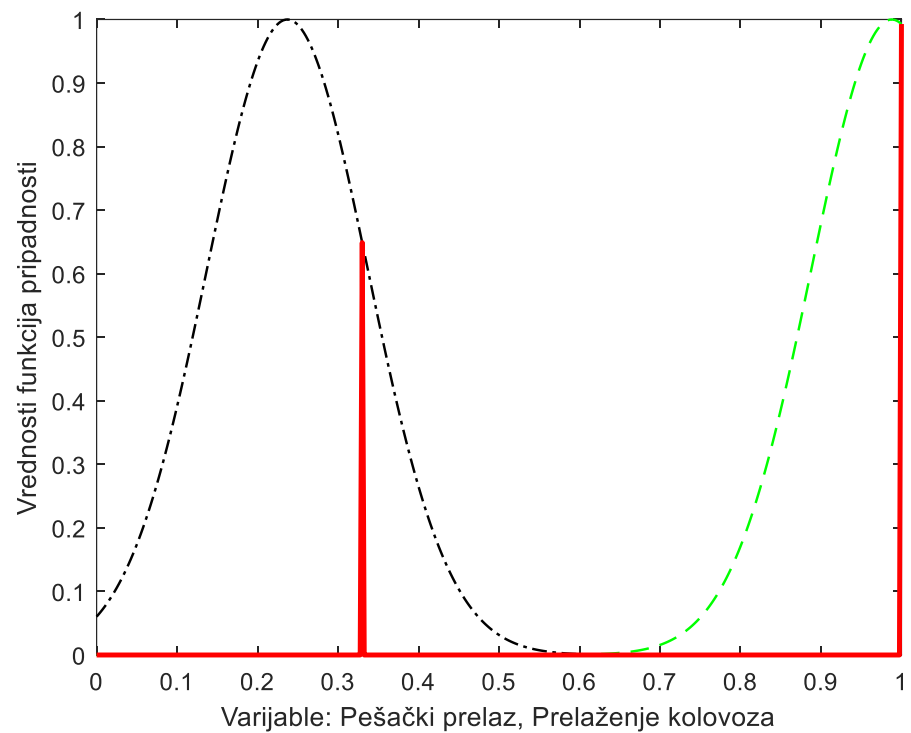
Kao što se vidi na Slici 6.28, za slučaj prvog TSK fazi pravila, funkcija pripadnosti varijable *Pešački prelaz* ima centar u 0.155, a funkcija pripadnosti varijable *Prelaženje kolovoza* ima centar u 0.994. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila je u vrednosti 0.216, dok kod varijable *Prelaženje kolovoza* presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila ima vrednost 0.998. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.28 crvenim stubićima.



Slika 6.28 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza

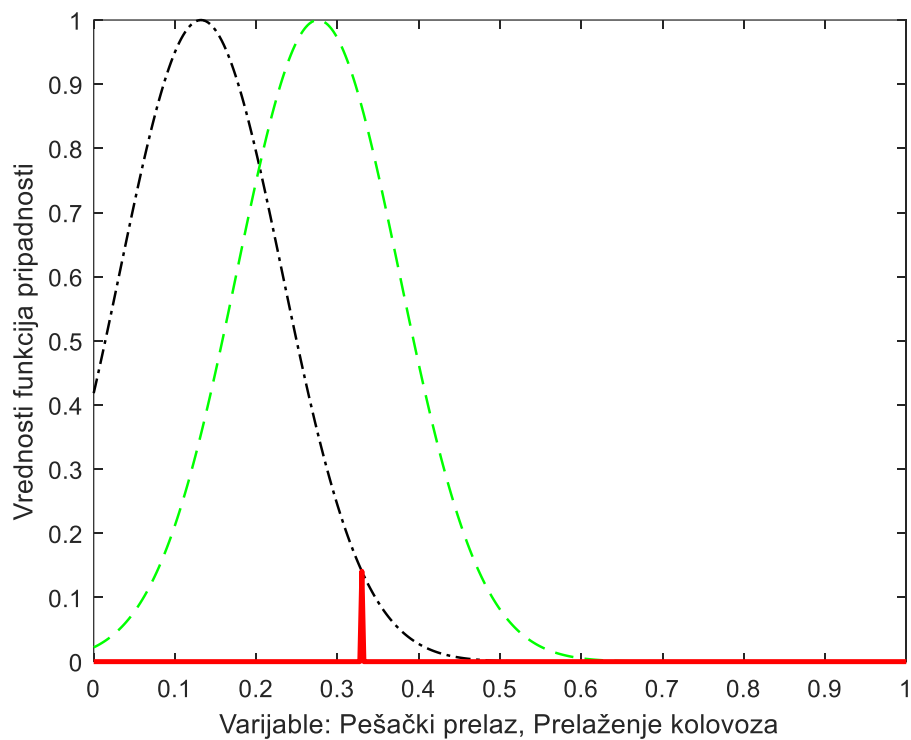
Kao što se vidi na Slici 6.29, za slučaj drugog TSK fazi pravila, funkcija pripadnosti za varijablu *Pešački prelaz* ima centar u 0.237, a funkcija pripadnosti za varijablu *Prelaženje kolovoza* ima centar u 0.988. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti drugog TSK fazi pravila ima vrednost 0.649 (u koordinati apscise 0), a za varijablu *Prelaženje kolovoza* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila u vrednosti 0.993. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.29 crvenim stubićima.





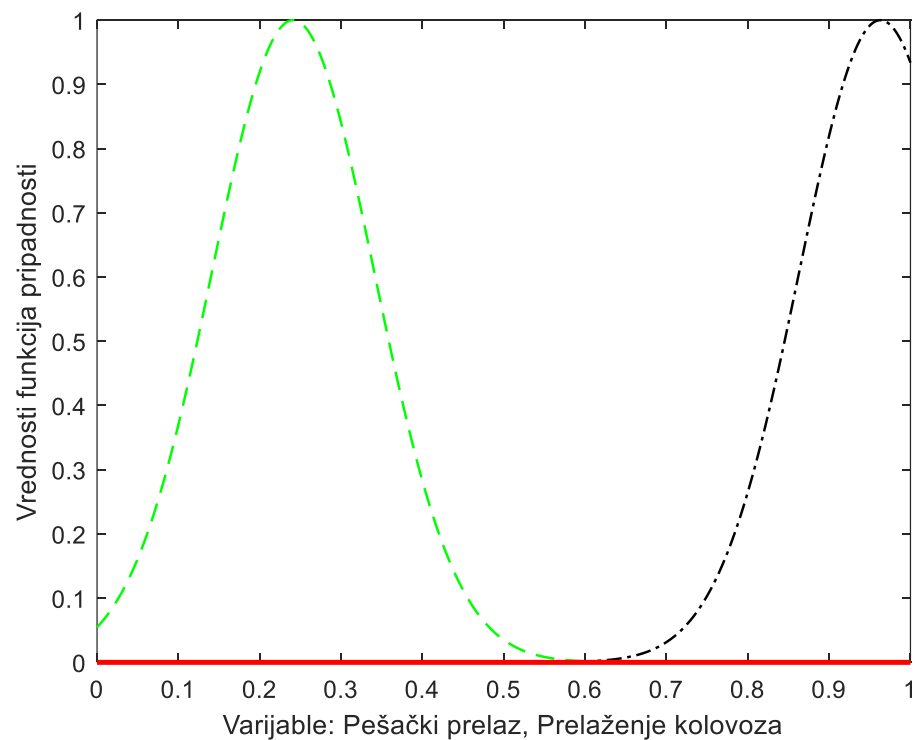
Slika 6.29 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza

Kao što se vidi na Slici 6.30, za slučaj trećeg TSK fazi pravila, funkcija pripadnosti za varijablu *Pešački prelaz* ima centar u 0.132, a funkcija pripadnosti za varijablu *Prelaženje kolovoza* ima centar u 0.276. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti trećeg TSK fazi pravila ima vrednost 0.141 (u koordinati apscise 0), a za varijablu *Prelaženje kolovoza* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila u vrednosti nula. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.30 crvenim stubićem.



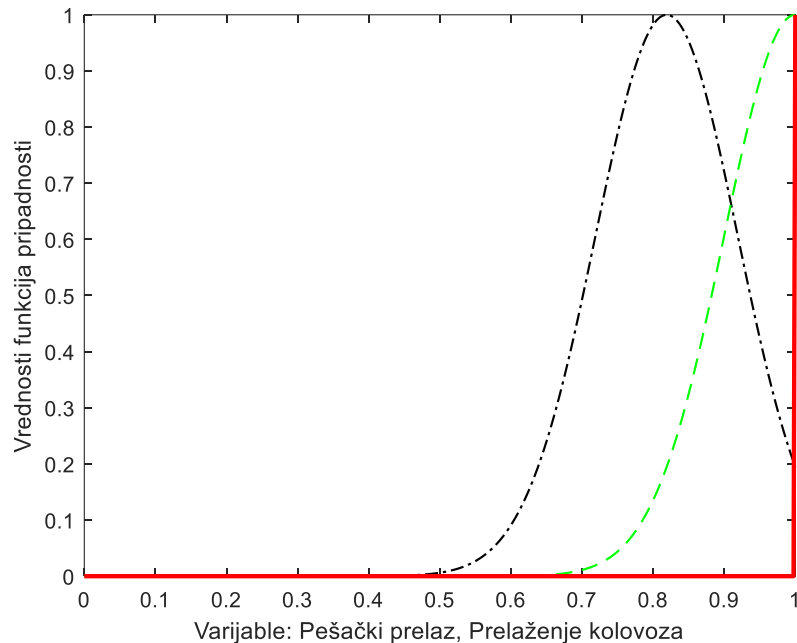
Slika 6.30 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza

Kao što se vidi na Slici 6.31, za slučaj četvrtog TSK fazi pravila, funkcija pripadnosti za varijablu *Pešački prelaz* ima centar u 0.963, a funkcija pripadnosti za varijablu *Prelaženje kolovoza* ima centar u 0.241. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti četvrtog TSK fazi pravila ima vrednost nula (nema preseka), a za varijablu *Prelaženje kolovoza* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila u vrednosti nula.



Slika 6.31 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza

Kao što se vidi na Slici 6.32, za slučaj petog TSK fazi pravila, funkcija pripadnosti za varijablu *Pešački prelaz* ima centar u 0.819, a funkcija pripadnosti za varijablu *Prelaženje kolovoza* ima centar u 1. U slučaju varijable *Pešački prelaz*, presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti petog TSK fazi pravila ima vrednost 0 (u koordinati apscise 0), a za varijablu *Prelaženje kolovoza* je presek egzaktne vrednosti asocijativnog pravila i Gausove funkcije pripadnosti prvog TSK fazi pravila u vrednosti 1. Preseci funkcija pripadnosti za obe varijable su označeni na Slici 6.32 crvenim stubićem.



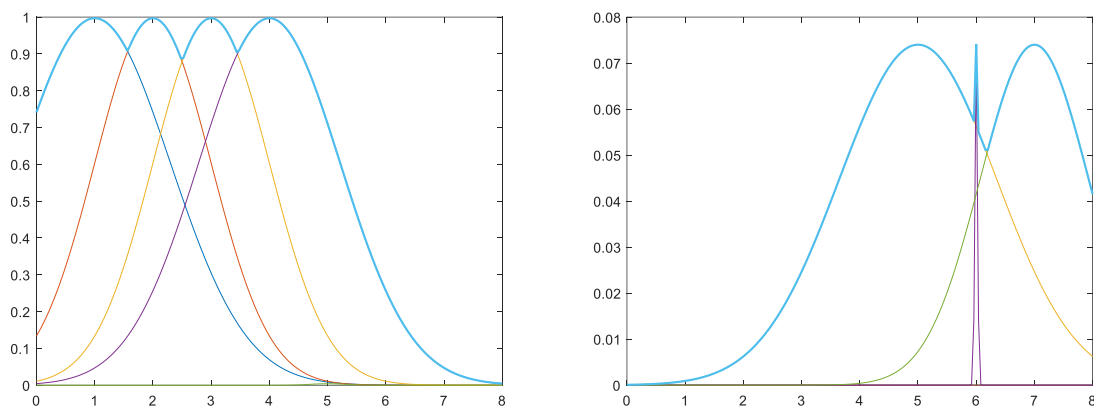
Slika 6.32 Presek uslovnih varijabli Pešački prelaz i Prelaženje kolovoza

Analiza pokrivenosti pojedinih TSK fazi pravila drugog modela sa asocijativnim pravilima je realizovana koristeći algoritam interpretiranja TSK fazi modela sa asocijativnim pravilima objašnjen u poglavlju 4.5. Dobijena je pokrivenost pravila drugog TSK fazi modela sa sedam selektovanih asocijativnih pravila. Postupak generisanja funkcija pokrivenosti fazi pravila sa asocijativnim pravilima kod drugog TSK fazi modela je ilustrovan na prvom i trećem TSK fazi pravilu, što je prikazano na Slici 6.33. Iz priloženih dijagrama je vidljiv način generisanja kumulativne funkcije pokrivenosti kombinovanjem više Gausovih funkcija preklapanja TSK fazi i asocijativnih pravila.

Pokrivenost kompletnog drugog TSK fazi modela sa asocijativnim pravilima je prikazana na Slici 6.34. Za prvo TSK fazi pravilo je lako uočiti da postoji preklapanje sa prva četiri asocijativna pravila. Visina funkcije preklapanja pokazuje nivo preklapanja posledičnih delova pravila (TSK fazi pravila sa asocijativnim pravilima) i u ovom slučaju ima vrednost blisku jedinici. Kod drugog TSK fazi pravila prisutno je manje preklapanje sa svim asocijativnim pravilima, izuzimajući šesto pravilo. Preklapanje posledičnih delova prva četiri asocijativna pravila (vrednost bliska 0.4), dok je kod petog i sedmog pravila vrednost oko 0.1. Kod trećeg TSK fazi pravila je slaba pokrivenost sa petim, šestim i sedmim asocijativnim pravilom sa maksimalnom vrednošću preklapanosti posledičnih delova pravila od 0.17.

Preklopljenost sa asocijativnim pravilima kod četvrtog TSK fazi pravila je vrlo slična kao kod prethodnog, s tim da preklopljenost posledičnih delova pravila iznosi ispod 0.07. Peto TSK fazi pravilo poseduje preklopljenost sa petim i sedmim asocijativnim pravilima, a intezitet preklopljenosti je blizak jedinici.

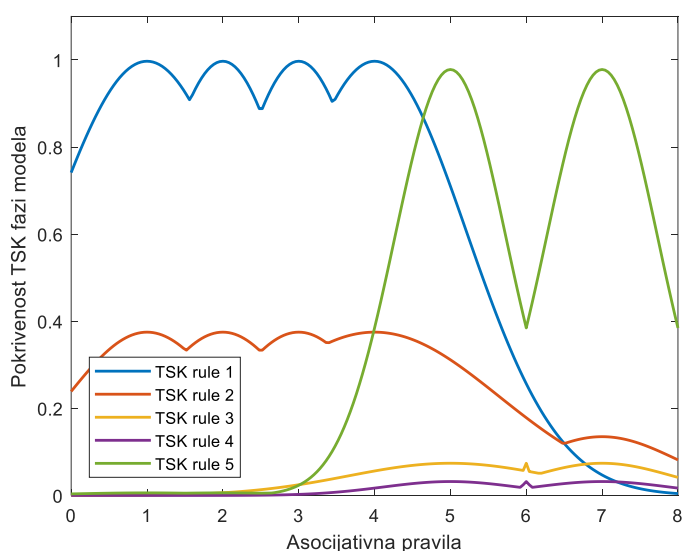
Veća širina funkcije kod sedmog asocijativnog pravila ilustruje veću pokrivenost ulaznog prostora nego kod petog asocijativnog pravila.



a) Generisanje funkcije pokrivenosti prvog fazi pravila drugog TSK fazi modela sa prva četiri asocijativna pravila

b) Generisanje funkcije pokrivenosti trećeg fazi pravila drugog TSK fazi modela sa asocijativnim pravilima 5, 6 i 7

Slika 6.33 Ilustracija postupka generisanja funkcija pokrivenosti fazi pravila sa asocijativnim pravilima kod drugog TSK fazi modela



Slika 6.34 Pokrivenost pravila drugog TSK fazi modela sa asocijativnim pravilima

Uspostavljena veza između asocijativnih i pravila TSK fazi modela omogućava interpretiranje fazi pravila drugog TSK fazi modela na sledeći način.

TSK fazi pravilo 1: Prvo pravilo predstavlja kombinaciju prva četiri asocijativna pravila. Imajući u vidu oblik i površinu ispod funkcije pokrivenosti za ovo pravilo, može se tvrditi da je ovo dominantno pravilo u drugom TSK fazi modelu. Posledica ovog pravila ima značenje „Bez povreda“. U uslovnom delu pravila dominira *Prelazak kolovoza* sa značenjem „Ostalo“. Pored toga, ovu varijablu karakteriše prelazak na ili izvan pešačkog prelaza.

TSK fazi pravilo 2: Ovo pravilo predstavlja kombinaciju prva četiri asocijativna pravila, zajedno sa sedmim pravilom. Posledični deo pravila ima dominantno značenje „Bez povreda“ i u mnogo manjem obimu značenje „Sa manjim povredama“. Nivo funkcije pokrivenosti drugog pravila pokazuje manji uticaj na ponašanje modela od prvog pravila. U uslovnom delu pravila dominira *Prelazak kolovoza* sa značenjem „Ostalo“. Pored toga, ovu varijablu karakteriše prelazak na ili izvan pešačkog prelaza, kao i prolazak kroz crveno svetlo kao oblik ponašanja.

TSK fazi pravilo 3: Pravilo predstavlja kombinaciju petog, šestog i sedmog asocijativnog pravila. Posledica ovog pravila ima značenje „Sa manjim povredama“, sa malim uticajem na ponašanje modela. U uslovnom delu pravila dominira varijabla *Ponašanje* sa vrednošću „Prolazak kroz crveno svetlo“, a prisutan je i uticaj prelaska kolovoza sa leve na desnu stranu.

TSK fazi pravilo 4: Dato pravilo takođe predstavlja kombinaciju petog, šestog i sedmog asocijativnog pravila. Posledica ovog pravila ima značenje „Sa manjim povredama“, ali sa još manjim uticajem na ponašanje modela od prethodnog pravila. U uslovnom delu pravila je slična situacija kao kod prethodnog pravila.

TSK fazi pravilo 5: Pravilo predstavlja kombinaciju asocijativnih pravila 5 i 7. Posledica ovog pravila ima značenje „Sa manjim povredama“, sa velikim uticajem na ponašanje drugog TSK modela. U uslovnom delu pravila dominira varijabla *Ponašanje* sa vrednošću „Prolazak kroz crveno svetlo“ i dodatno prelazak na pešačkom prelazu.

Na kraju, kao i za prvi TSK fazi model, može se zaključiti da postoje slaganja dobijenih rezultata dobijenih na osnovu razmatrane dve varijante algoritma: interpretera TSK fazi modela sa asocijativnim pravilima i interpretera asocijativnih pravila sa TSK fazi pravilom. Poređenjem podataka sa Slike 6.34 i Slike 6.22 može se

uočiti da su kod prvih pet asocijativnih pravila najveće sličnosti vezane za prvo TSK fazi pravilo. Posmatrajući Sliku 6.34, ovo pravilo je opisano Gausovom krivom koja je najviša po intezitetu (blizu 1), a prema jačini zavisnosti sa asocijativnim pravilima se javljaju redom asocijativna pravila 2 i 3, a potom i pravila 1 i 4, kao što se vidi na Slici 6.22. Iako je Gausova kriva drugog TSK fazi pravila niža po intezitetu, ona pokriva najveći prostor razmatranja (zauzima gotovo ceo opseg  $[0,1]$ ) i može se primetiti da je za ovo TSK fazi pravilo vezan najveći broj asocijativnih pravila (skoro svih). Na osnovu Slike 6.34 se može zaključiti da su najdominantnija pravila redom: prvo TSK fazi pravilo koje pokriva udese bez povreda, peto TSK fazi pravilo koje je isključivo vezano za posledice sa manjim povredama i drugo TSK fazi pravilo koje uključuje oba slučaja posledica.

#### 6.4 Diskusija rezultata za oba TSK fazi modela

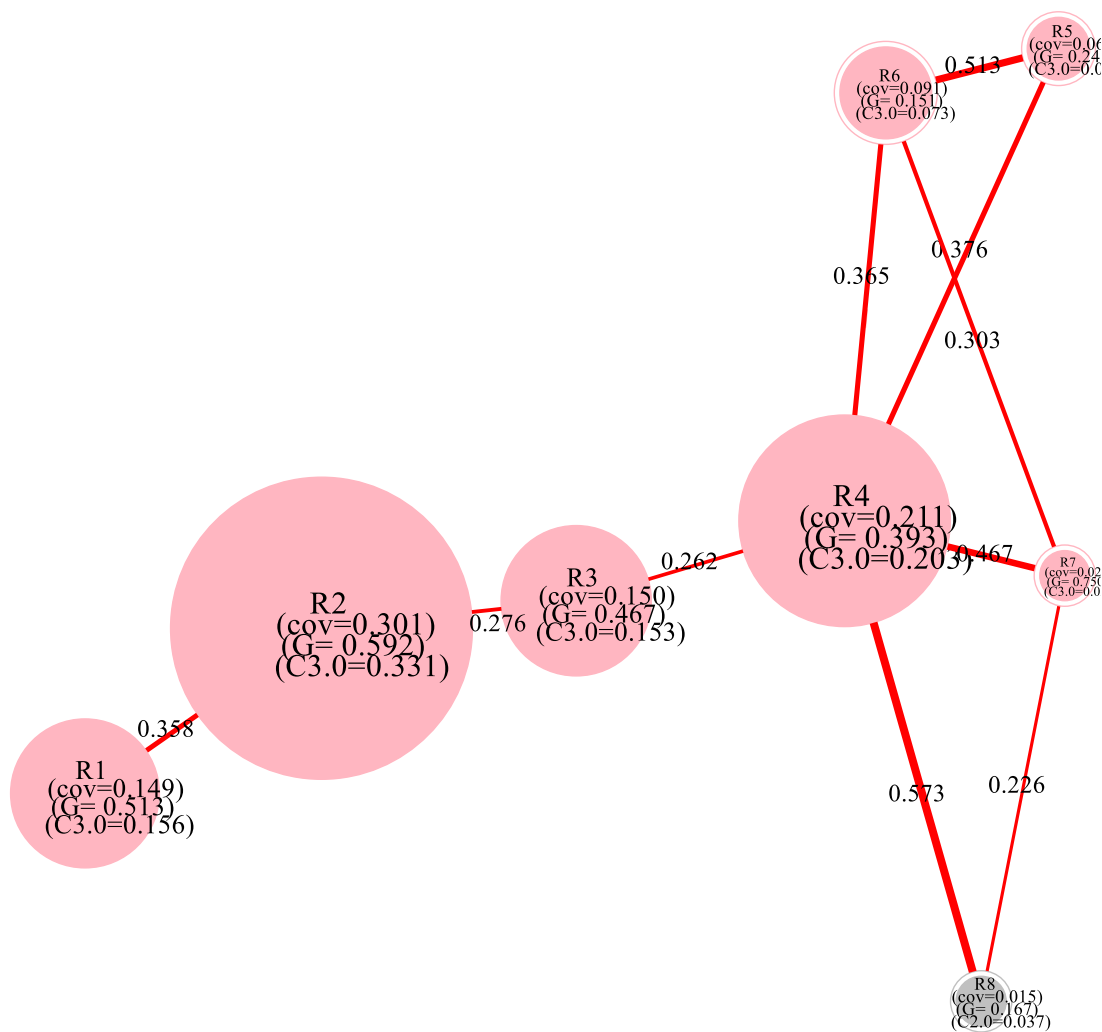
Slike 6.18 i 6.34 jasno ilustruju asocijaciju TSK fazi pravila sa pojedinim asocijativnim pravilima. Prvo je lako uočiti koja asocijativna pravila su povezana sa određenim TSK fazi pravilom. Pri tome, imajući u vidu posledice asocijativnih pravila, lako se može zaključiti koji tip posledica svako TSK fazi pravilo sadrži. S obzirom da je visina karakteristike nekog TSK fazi pravila određena intenzitetom preklapanja posledičnih delova pravila, može se odrediti uticaj pojedinih pravila u modelu. Širina karakteristike pokazuje koliko dato TSK fazi pravilo široko pokriva prostor problema. Širina pokrivenosti je funkcija nivoa preklapanja varijabli uslovnih delova pravila, broja takvih varijabli i broja pravila kod kojih su se takva preklapanja našla.

Na primeru prvog TSK fazi modela, prva dva pravila odgovaraju posledicama „Sa manjim povredama“, dok je treće određeno značenjem „Bez povreda“. Sa slike 6.18 je jasno uočljivo da je treće pravilo dominantno u modelu, jer će po visini karakteristike, a posebno po površini koju karakteristika pokriva, ovo pravilo značajno uticati na krajnji izlaz TSK fazi modela.

Za slučaj drugog TSK fazi modela, prvo pravilo odgovara posledicama „Bez povreda“, dok pravila 3, 4 i 5 određuju posledice sa „Manjim povredama“. Drugo pravilo dominantno određuje posledice „Bez povreda“, ali u manjoj meri i posledice sa „Manjim povredama“. Dominantno pravilo u ovom TSK fazi modelu je prvo pravilo, međutim peto i drugo pravilo imaju značajnu ulogu. Međutim, drugo pravilo je interesantno zbog veze sa asocijativnim pravilima koja imaju obe vrste posledica: „Manje povrede“ i „Bez povreda“.

## 6.5 Rezultati vizualizacije - Fingrams

Kao što je već objašnjeno, *Fingrams* (*Fuzzy Inference-grams*) grafički predstavljaju fazi modele koji se zasnivaju na pravilima [97, 98]. Slika 6.35 prikazuje izgled *Fingrams* grafa kreiranog za Wang-Mendel fazi model (*GUAJE-WM*) generisanog na istim skupom uzoraka. *GUAJE-WM* graf prikazuje Wang-Mendel fazi model sa 8 pravila (*Fingrams* čvorova) sa najznačajnijim relacijama (vezama) koje postoje između njih. Čvorovi sadrže identifikator pravila ( $R_i$ ), pokrivenost datog pravila ( $cov$ ), dobrotu pravila ( $G$ ), tj. kako se pravilo ponaša s obzirom na dostupne uzorke problema, i relativnu pokrivenost klase izlaza pravila ( $CR_i$ ) [97]. Pokrivenost  $CR_i$  se računa kao broj uzoraka koji su pokriveni pravilom koje pripada datoj klasi podeljen ukupnim brojem uzoraka iste klase.

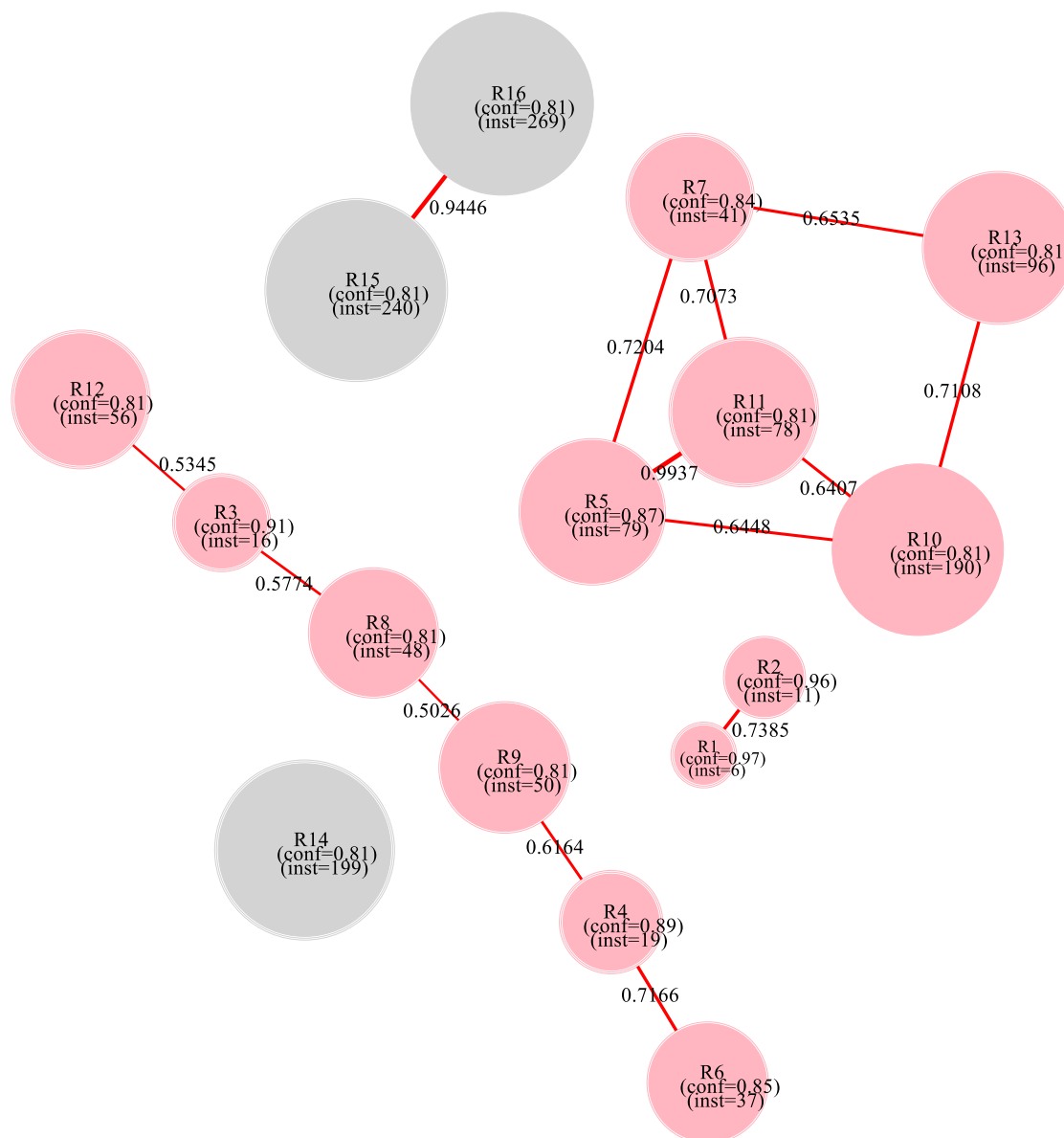


Slika 6.35 *Fingrams* graf *GUAJE WM* fazi modela



Sa Slike 6.35 se jasno vidi da model dominantno pokriva klasu *Lakše povrede* i da moguća nedoslednost postoji između pravila  $R_4$  i  $R_8$ , jer su u direktnoj vezi, a imaju različite posledice, odnosno pripadaju različitim klasama. Pravilo  $R_4$  je dominantno pravilo ovog modela, s obzirom na to da ima najviše veza i smešteno je u središtu *Fingrams* grafa, dok je pravilo  $R_2$  pokriveno sa najviše uzoraka.

Slika 6.36 prikazuje *Fingrams* graf modela sa 16 asocijativnih pravila (*Fingrams* čvorova) i najvažnije relacije (veze) koje postoje među njima. Veze sa težinama ispod predefinisano praga sa vrednošću 0.5 su izostavljene. Čvorovi *Fingrams* grafa sadrže kao attribute identifikator pravila  $R_i$ , nivo poverenja (*conf*) i broj uzoraka koji su pokriveni datim pravilom (*inst*). Usled velikih razlika u broju uzoraka koji su pokriveni sa pravilima, veličina čvora je načinjena u proporciji sa logaritmom pridruženog broja uzoraka datom čvoru.



Slika 6.36 Fingrams graf modela asocijativnih pravila

Sa Fingrams grafa na Slici 6.36 se može lako uočiti odnos pravila koji pripada različitim klasama. Čvorovi R<sub>14</sub>, R<sub>15</sub> i R<sub>16</sub> pripadaju klasi – *Bez povreda* (čvorovi sive boje), dok ostali čvorovi pripadaju klasi – *Lake povrede* (čvorovi crvene boje). Parovi pravila (R<sub>1</sub>, R<sub>2</sub>), (R<sub>15</sub>, R<sub>16</sub>) i (R<sub>5</sub>, R<sub>11</sub>) ukazuju na redundantna pravila i moguće kandidate za isključivanje iz modela usled velikog udela zajedničkih uzoraka. Međutim, kod pravila R<sub>15</sub> i R<sub>16</sub>, pored toga što su slična, karakteristično je da nisu povezana sa ostalim čvorovima što ukazuje na to da su ova pravila drugačija od ostalih jer ih pokrivaju uzorci koje druga pravila ne pokrivaju. Pravilo R<sub>14</sub> je važno za klasu – *Bez povreda*, dok klasa – *Lake povrede* poseduje više pravila koja dobro predstavljaju ovu klasu, uključujući pravila povezana sa pravilima R<sub>12</sub> i R<sub>10</sub>.

## 6.6 Komparativna analiza modela interpretabilnosti zasnovanih na pravilima

Kako bi se postigla objektivnija procena razmatrane koherentnosti modela zasnovanih na pravilima, korišćeno je nekoliko objektivnih mera interpretabilnosti koje obezbeđuju procenu složenosti na osnovu parametara modela (broj pravila, uslova, promenljivih, fazi particija, i sl.). U daljem tekstu su ukratko opisane korišćene standardne mere interpretabilnosti.

**Ukupna dužina pravila, UDP** (eng. *Total rule length, TRL*): Predstavlja broj uslova koji su definisani za sve ulaze (NI) i koji su korišćeni u svim pravilima (NR):

$$TRL = \sum_{i=1}^{NR} \sum_{j=1}^{NI} c(P_{ij}), \quad (6.1)$$

gde je  $c(P_{ij})$  preduslov za  $j$ -ti ulaz u  $i$ -tom pravilu i koji je jednak vrednosti 1 ukoliko se pojavljuje u datom pravilu, a u suprotnom je jednak 0.

**Prosečna dužina pravila** (eng. *Average rule length*): Ukupna dužina pravila podeljena sa brojem pravila.

**Akumulirana složenost pravila, ASP** (eng. *Accumulated rule complexity, ARC*) [103]: Računa se kao zbir kompleksnosti pravila izmerena za svih NR pravila. Složenost pravila se meri kao proizvod svih  $p$ -složenosti preduslova za NI ulaza koji se koriste u pravilu. Indeks  $p$ -složenosti se računa uzimanjem u obzir broja elementarnih izraza definisanih za svaku ulaznu vrednost  $NL_j$  i broja elementarnih lingvističkih pojmova koji učestvuju u pravilu. Ukoliko lingvistički pojmovi sadrže jedino elementarne izraze, akumulirana složenost pravila ASP se računa kao:

$$ASP = \sum_{i=1}^{NR} \left[ \prod_{j=1}^{NI} \left( 2 - \frac{1}{NL_j} \right) \right] \quad (6.2)$$

**Teoretski izvršena pravila, TIP** (eng. *Theoretical fired rules, TFR*) [103]: Daje procenu prosečnog broja pravila koja mogu biti izvršena.

$$TIP = \frac{1}{\sum_{i=1}^{NR} \sum_{j=1}^{NR} [fired(R_i, R_j)]},$$

$$fired(R_i, R_j) = \begin{cases} 1, & i \neq j \wedge \text{oba pravila se preklapaju} \\ 0, & \text{u suprotnom} \end{cases} \quad (6.3)$$

gde termin  $fired()$  definiše istovremeno izvršena pravila. Za par pravila ( $R_i, R_j$ ) se smatra da bi se mogla istovremeno izvršiti kada se svi njihovi lingvistički izrazi preklapaju u potpunosti ili bar delimično.

**Naukov indeks** (eng. *Nauck's Index*) [104]: Naukov indeks meri interpretabilnost klasifikacionih fazi modela, u smislu složenosti (*comp*), nivoa pokrivenosti fazi particija ( $\overline{cov}$ ) i složenosti fazi particija ( $\overline{part}$ ).

$$\text{Naukov Indeks} = comp \cdot \overline{cov} \cdot \overline{part} \quad (6.4)$$

Mera složenosti (*comp*) predstavlja složenost klasifikatora izračunatu kao broj klasa *m* podeljen sa ukupnim brojem uslova u svim pravilima.

$$comp = m / \sum_{i=1}^{NR} n_i \quad (6.5)$$

Druga komponenta,  $\overline{part}$ , označava prosečan normalizovan indeks particije za sve ulazne promenljive. Ova komponenta daje veću težinu particiji sa visokom granularnošću i računa se kao recipročna vrednost broja lingvističkih pojmova  $p_j$  minus jedan, za svaku *j*-tu ulaznu promenljivu.

$$\overline{part} = \frac{1}{NI} \sum_{j=1}^{NI} \frac{1}{p_j - 1} \quad (6.6)$$

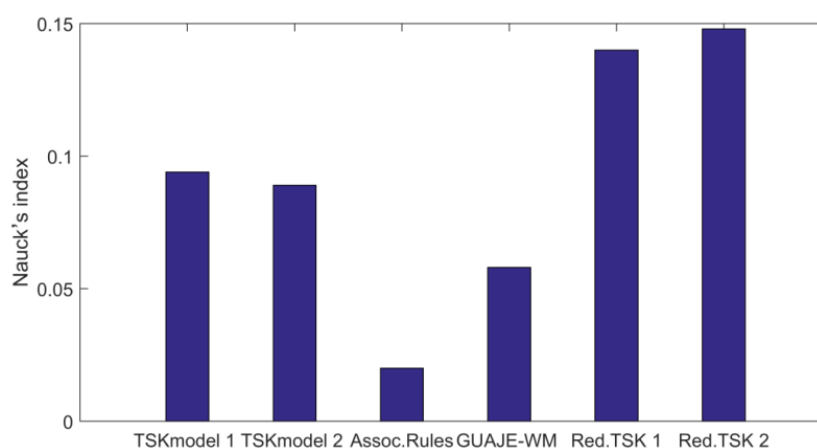
Treća komponenta Naukovog indeksa, nivo pokrivenosti fazi particija - indeks pokrivenosti ( $\overline{cov}$ ), je prosečan stepen normalizovane pokrivenosti fazi particija. Takođe, ova komponenta ima vrednost jednaku jedinici za jake fazi particije.

Prethodno opisane mere interpretabilnosti se koriste za poređenje oba kreirana TSK fazi modela, a pogotovo njihove verzije sa redukovanim brojem pravila. Na osnovu predloženog TSK-AR algoritma Povezivača, pravilo broj 2 je uklonjeno iz prvog TSK fazi modela, dok pravila 4 i 5 nisu razmatrana za drugi TSK fazi model. Dodatno, u cilju komparacije, razmatrana su još dva modela (videti Tabelu 6.9). Naime, prvo su izračunate mere interpretabilnosti za model sa asocijativnim pravilima, gde su razmatrana samo pravila koja imaju viši nivo poverenja od 0.75 vrednosti. Drugi model je generisan pomoću softverskog alata *GUAJE* (eng. *Generating Understandable and Accurate fuzzy models in a Java Environment*) [103, 105, 106]. *GUAJE* je alat za razvoj interpretabilnih i preciznih fazi sistema kombinovanjem nekoliko postojećih alata sa slobodno raspoloživim izvornim kodom (eng. *open source*). U cilju indukcije dodatnog modela sa fazi pravilima je korišćen je Wang - Mendelov (eng. *Wang and Mendel's*) postupak koji je ugrađen u alat *GUAJE*. To je jednostavan i dobro poznat algoritam generisanja fazi pravila (*GUAJE-WM*) [92].

Tabela 6.9 Izračunate kvantitativne mere interpretabilnosti modela zasnovanih na pravilima

Mere interpretabilnosti	TSK model 1	TSK model 2	Asocijativna pravila	GUAJE-WM	Redukovan TSK model 1	Redukovan TSK model 2
Broj pravila	3	5	16	8	2	3
Ukupna dužina pravila	15	15	36	12	10	9
Prosečna dužina pravila	5	3	2.25	1.5	5	3
ASP (ARC)	27.083	48.75	57.34	19.8	16.25	32.5
TIP (TFR) (prosek)	12	10	1.375	3	8	6
Naukov indeks	0.094	0.089	0.02	0.058	0.14	0.148

Oba redukovana TSK fazi modela se odlikuju mnogo manjim **Brojem pravila** i **Ukupnom dužinom pravila** u poređenju sa ostalim modelima. Manji **Broj pravila** i manja **Ukupna dužina pravila** poboljšavaju interpretabilnost modela. Asocijativna pravila su dala bolju interpretabilnost nego ostali modeli u skladu sa prosečnom TIP vrednošću, ali ovo je i bilo očekivano s obzirom da su svi ostali modeli karakterisani izvršavanjem fazi pravila. *GUAJE-WM* model je pokazao najbolje karakteristike za **Prosečnu dužinu pravila**, s obzirom da procenjuje modele na osnovu najjednostavnijih pravila. Međutim, najpotpuniji indeks interpretabilnosti među izloženim merama je Naukov indeks. Model je interpretabilniji kada Naukov indeks ima veću vrednost. Slika 6.37 ilustruje značajno veću interpretabilnost kod redukovanih TSK fazi modela u poređenju sa ostalim razmatranim modelima.



Slika 6.37 Vrednosti Naukov-og indeksa razmatranih modela zasnovanih na pravilima



## 7. Zaključak

Predmet istraživanja disertacije je proučavanje razumljivosti ponašanja Takagi-Sugeno-Kang (TSK) fazi logičkih modela i rad na njegovom poboljšanju. Naime, TSK fazi logički model obično odlikuje visoka tačnost, ali i nedovoljna razumljivost. Rezultat istraživanja je nova metodologija i radni okvir za poboljšanje razumljivosti ponašanja TSK fazi modela. Razvijena metodologija poboljšava razumljivost TSK fazi modela pronalaženjem podudarnosti između fazi i asocijativnih pravila. Kriterijum njihovog podudaranja su nivoi preklapanja u vrednostima promenljivih koje se pojavljuju u uslovnom i posledičnom delu fazi i asocijativnih pravila. Predložena metodologija obezbeđuje vizualnu predstavu najvažnijih pravila i jasno oslikava vezu između određenih TSK fazi i asocijativnih pravila.

Shodno postavljenoj hipotezi H1 vezanoj za postupak selekcije obeležja u okviru razvijene metodologije pokazano je da specifična realizacija hibridnog algoritma optimizacije jata rešenja sa veštačkim neuronskim mrežama (PSO+ANN) omogućava izbor nezavisnih ulaznih veličina neophodnih za generisanje fazi logičkog modela. Specifičnost realizacije PSO+ANN algoritma je u tome da: (a) postupak čini okruženje koje po PSO obrascu generiše veliki broj podskupova varijabli kao kandidat liste varijabli; (b) korišćenju neuronske mreže kao fitness funkcije i (c) ceo postupak se višestruko ponavlja kako bi se smanjila stohastička karakteristika PSO+ANN algoritma i lakše uočile najfrekventnije izdvojene varijable. Konačno selektovani podskup varijabli sa kojim su generisani modeli zasnovani na pravilima je pokazao da omogućava razumevanje razmatranog problema.

Razmatranje činjenice izložene u okviru hipoteze H2 pokazuje da je projektovanim postupkom za povećavanje razumljivosti modela zasnovanog na pravilima moguće izdvajanje najznačajnijih pravila. Značajnost jednog fazi pravila je proporcionalna visini i širini kumulativne funkcije pokrivenosti pravila. Drugim rečima, što je intenzitet preklapanja pravila posledičnih delova pravila viši i broj uključenih pravila veći, fazi pravilo je važnije. Pored toga, izračunate objektivne mere interpretabilnosti, uključujući i Naukov indeks, su jasno pokazale da se razumljivost ponašanja TSK fazi modela znatno povećala kada su u modelu preostala samo dominantna pravila.

Polazeći od činjenice da Takagi-Sugeno-Kang forma fazi pravila nije dovoljno razumljiva, hipotezom H3 se pretpostavila mogućnost razvoja postupka za povećavanje njihove razumljivosti. Istraživanje je pokazalo da je moguće razviti postupak koji omogućava transformaciju fazi pravila Takagi-Sugeno-Kang forme u pravila jasnije i uobičajene AKO-ONDA forme. Razvijeni postupak koristi kao svoj ulaz, pored parametara TSK fazi modela, i skup asocijativnih pravila generisan nad istim skupom podataka kao i razmatrani TSK fazi model. Asocijativna pravila čije se vrednosti atributa preklapaju sa vrednostima ulazno-izlaznih varijabli razmatranog fazi pravila, na način definisan metodologijom, povezuju se sa datim fazi pravilom. Podskupovi povezanih asocijativnih pravila sa svakim od TSK fazi pravila omogućavaju lakše razumevanje fazi pravila, procenu uticaja pojedinih varijabli, kao i ocenu važnosti pravila u ponašanju fazi modela.

Nesumnjivo najočigledniji način povećavanja razumljivosti modela zasnovanog na fazi pravilima omogućava vizualizacija TSK fazi modela. Ova pretpostavka navedena u hipotezi H4 je takođe realizovana u okviru metodologije razvijene u disertaciji. Grafički prikaz koji predstavlja rezultat razvijene metodologije jednostavno vizualizuje pronađene veze između asocijativnih i TSK fazi pravila. Naime, grafički prikaz omogućava jasan uvid u postojeće sličnosti između TSK fazi i asocijativnih pravila. Vizualizacija modela je predstavljena u obliku dijagrama u kojem apscisa predstavlja prostor razmatranja problema opisan sa podskupom izdvojenih asocijativnih pravila, dok ordinata opisuje intenzitet funkcije pokrivenosti TSK modela sa povezanim asocijativnim pravilima. Funkcija pokrivenosti koja opisuje vezu između datog fazi pravila i povezanih asocijativnih pravila svojom formom vrlo očigledno pokazuje da: (a) visina karakteristike (funkcije pokrivenosti) određuje uticaj pojedinih pravila u modelu, a zavisi od intenziteta preklapanja posledičnih delova pravila i (b) širina karakteristike pokazuje koliko dato TSK fazi pravilo široko pokriva prostor problema, pri čemu je širina pokrivenosti funkcija nivoa preklapanja varijabli uslovnih delova pravila, broja takvih varijabli i broja pravila kod kojih su se takva preklapanja našla.

U okviru disertacije su realizovana dva TSK fazi modela u cilju procene uticaja broja pravila i parametara TSK fazi modela na njihovu razumljivost. Prvi generisani TSK fazi model sadrži pet ulaza i ima tri pravila. Drugi TSK fazi model ima tri ulaza i sastoji se od pet pravila. Zahvaljujući ovako odabranim konfiguracijama, oba modela imaju približno jednak broj parametara. Oba TSK fazi modela su generisana na podacima o saobraćajnim nezgodama u gradskoj sredini sa učešćem



pešaka. Analiza rezultata razvijene metodologije je pokazala da je moguće prepoznati situacije i uzroke saobraćajnih nezgoda koji su navedeni u daljem tekstu.

U prvom TSK fazi modelu je očigledno da je treće pravilo dominantno i da opisuje slučajeve saobraćajnih nezgoda bez povreda. Prva dva pravila modela odgovaraju saobraćajnim nezgodama sa manjim povredama. U drugom TSK fazi modelu je dominantno prvo pravilo koje opisuje saobraćajne nezgode bez povreda. Peto i drugo pravilo imaju značajnu ulogu u drugom TSK fazi modelu. Pri tome, peto fazi pravilo odgovara slučajevima saobraćajnih nezgoda sa manjim povredama, dok je drugo fazi pravilo posebno interesantno jer odgovara saobraćajnim nezgodama sa manjim povredama i bez povreda.

Analizom uzročno-posledičnih relacija u razmatranim pravilima oba TSK fazi modela može se zaključiti sledeće: a) U slučaju saobraćajnih nezgoda bez povreda postoji snažna veza sa prelaskom kolovoza na ili izvan pešačkog prelaza. Primećena je veza slabijeg inteziteta između saobraćajnih nezgoda bez povreda i osoba muškog pola, kao i osoba starosti od 30 do 64 godine; b) U slučaju saobraćajnih nezgoda sa manjim povredama postoji snažna veza sa prelaskom kolovoza na crveno. Slabija veza je detektovana i sa osobama ženskog pola, kao i sa osobama starosti između 15 i 29 godina.

Savremene tehnike mašinskog učenja kao što su 'duboko' učenje ili učenje više kombinovanih modela, tkz. učenje ansambla modela, odlikuje ogromna količina podataka raspoloživih za obuku i veliki broj parametara modela koje treba podesiti. Postoje različita istraživanja sa zadatkom da omoguće bolju razumljivost ovih inače obično vrlo tačnih prediktora ili klasifikatora. Postojeća istraživanja se zasnivaju na ideji da se, nakon generisanja složenog i vrlo tačnog modela gde se ne vodi računa o njegovoj razumljivosti, pristupi izradi naknadnog modela koristeći druge metode predstavljanja znanja, uključujući npr. stabla ili liste odlučivanja. Principi rada razvijenog okvira za poboljšanje razumljivosti TSK fazi modela bi se mogli prilagoditi i iskoristiti na modelima 'dubokog' učenja ili ansambla modela. U ovom slučaju bi zadatak bio da se putem asocijativnih i fazi pravila objasne složene i skrivene relacije između ulaznih i izlaznih veličina modela, a koje su ugrađene u strukturu i parametre ovih modela tokom njihovog učenja. Pri tome je vrlo važno da se pronađu najvažnije ulazne veličine i dominantna pravila koja određuju ponašanje modela u različitim slučajevima.



## 8. Literatura

- [1] L. A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Transactions on Systems, Man and Cybernetics* 3, pp.28–44, 1973.
- [2] U. Bodenhofer, P. Bauer, A formal model of interpretability of linguistic variables, pp. 524–545, 2003.
- [3] J. Casillas, O. Cordón, F. Herrera, L. Magdalena, Accuracy improvements in linguistic fuzzy modeling, *Studies in Fuzziness and Soft Computing*, vol. 129, Springer, 2003.
- [4] J. Casillas, O. Cordón, F. Herrera, L. Magdalena, Interpretability Issues in Fuzzy Modeling, *Studies in Fuzziness and Soft Computing*, vol. 128, Springer-Verlag, 2003.
- [5] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, Interpretability Improvements to Find the Balance Interpretability-Accuracy in Fuzzy Modeling: An Overview, *Interpret. Issues Fuzzy Model.*, vol. 128, pp. 3–22, 2003.
- [6] Z. C Lipton, The mythos of model interpretability, *Communications of the ACM*, Vol. 61 (10), pp. 36-43, October 2018.
- [7] P. Liang. Provenance and contracts in machine learning. In *Proceedings of the 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*, 2016.
- [8] A. Adoko, C. Gokceoglu, L. Wu, and Q. Zuo, Knowledge-based and data-driven fuzzy modeling for rockburst prediction, *Int. J. Rock Mech. Min. Sci.*, vol. 61, pp. 86–95, 2013.
- [9] A. Adoko, C. Gokceoglu, L. Wu, and Q. Zuo, Knowledge-based and data-driven fuzzy modeling for rockburst prediction, *Int. J. Rock Mech. Min. Sci.*, vol. 61, pp. 86–95, 2013.
- [10] S. Guillaume, B. Charnomordic, Generating an interpretable family of fuzzy partitions from data, *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 3, pp. 324–335, 2004.
- [11] R. Alcalá, J. Casillas, O. Cordón, A. González, F. Herrera, A genetic rule weighting and selection process for fuzzy control of heating, ventilating and air conditioning systems, *Eng. Appl. Artif. Intell.*, vol. 18, no. 3, pp. 279–296, 2005.

- [12] A. Rüstern, Transparent Fuzzy Systems in Modelling and Control, Interpret. Issues Fuzzy Model, pp. 452–476, 2003.
- [13] C. C. Lee, Fuzzy logic in control systems: fuzzy logic controller- part I and II, IEEE Trans Syst Man Cybern, 20(2):404–435, 1990.
- [14] D. Kukulj, Design of adaptive Takagi–Sugeno–Kang fuzzy models. Applied Soft Computing, Vol 2, No. 2, pp. 89–103, December 2002. [https://doi.org/10.1016/S1568-4946\(02\)00032-7](https://doi.org/10.1016/S1568-4946(02)00032-7)
- [15] S. Haykin, Neural networks and learning machines, Pearson Education, Upper Saddle River, NJ, Third edition, 2009.
- [16] L. Magdalena, Designing interpretable Hierarchical Fuzzy Systems, IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE), pp 1–8, 2018. doi: <https://doi.org/10.1109/FUZZ-IEEE.2018.8491452>.
- [17] E. H. Mamdani, Application of Fuzzy Algorithm for Control of Simple Dynamic Plant, Proc. IEEE, 121, 1585-1888, 1974.
- [18] F. Yang, M. Du and X. Hu, Evaluating Explanation Without Ground Truth in Interpretable Machine Learning, *ArXiv abs/1907.06831* (2019).
- [19] T. Zhou, H. Ishibuchi, S. Wang, Stacked-structure-based hierarchical Takagi-Sugeno-Kang fuzzy classification through feature augmentation. IEEE Trans Emerging Top Comput Intell 1(6):421–436, 2017. DOI: 10.1109/TETCI.2017.2761915.
- [20] Y. Zhang, H. Ishibuchi, S. Wang, Deep Takagi–Sugeno–Kang fuzzy classifier with shared linguistic fuzzy rules. IEEE Trans Fuzzy Syst, 26(3):1535–1549, 2018. <https://doi.org/10.1109/TFUZZ.2017.2729507>.
- [21] A. Riid, E. Rüstern, Identification of transparent, compact, accurate and reliable linguistic fuzzy models. Inf Sci, 181(20):4378–4393, 2011. <https://doi.org/10.1016/j.ins.2011.01.041>.
- [22] W. L. Ho, W. L. Tong, C. Quek, An evolving Mamdani-Takagi-Sugeno based neural-fuzzy inference system with improved interpretability-accuracy, FUZZ-IEEE, 2010. <https://doi.org/10.1109/FUZZY.2010.5584831>.
- [23] S. M. Zhou, J. Q. Gan, Extracting Takagi-Sugeno fuzzy rules with interpretable submodels via regularization of linguistic modifiers. IEEE Trans Knowl Data Eng, 21(8):1191–1204, 2009. <https://doi.org/10.1109/TKDE.2008.208>.
- [24] B. Qin, B., F.-L. Chung, S. Wang, Biologically Plausible Fuzzy-Knowledge-Out and Its Induced Wide Learning of Interpretable TSK Fuzzy Classifiers. IEEE Trans. on Fuzzy Systems, 1–1, 2019. doi:10.1109/tfuzz.2019.2907497.

- [25] H. Haider, N. A. Mat Isa, M. M. Ahmed, A Structural Evolving Approach for Fuzzy Systems. *IEEE Transactions on Fuzzy Systems*, 1–1. 2019, doi:10.1109/tfuzz.2019.2904928.
- [26] A. G. Casado, P. C. Marchal, J. G. Ortega, J. G. Garcia, Visualization and Interpretation Tool for Expert Systems based on Fuzzy Cognitive Maps. *IEEE Access*, vol.7, pp.6140-6150, 2018. doi:10.1109/access.2018.2887355.
- [27] Z. Xi, G. Panoutsos, Interpretable Machine Learning: Convolutional Neural Networks with RBF Fuzzy Logic Classification Rules, *Int. Conference on Intelligent Systems (IS)*, pp.448-454, 2018. doi:10.1109/is.2018.8710470.
- [28] Y. Zhang, J. Dong, J. Zhu, C. Wu, Common and Special Knowledge-Driven TSK Fuzzy System and Its Modeling and Application for Epileptic EEG Signals Recognition, *IEEE Access*, vol.7, pp.127600-127614, 2019. doi:10.1109/ACCESS.2019.2937657.
- [29] Y. Zhang, X. Li, J. Zhu, C. Wu, Epileptic EEG Signals Recognition Using a Deep View-Reduction TSK Fuzzy System with High Interpretability, *IEEE Access*, vol.7, pp. 137344- 137354, 2019, doi:10.1109/ACCESS.2019.2942641.
- [30] X. Meng, Q. Hou, Y. Shi, Research on Accident Prediction Models for Freeways in Mountainous and Rolling Areas, *Seventh International Conference on Measuring Technology and Mechatronics Automation*, 2015, pp.838-838.
- [31] Y. Rujun, L. Xiuqing, Study on traffic accidents prediction model based on RBF neural network, *2nd International Conference on Information Engineering and Computer Science (ICIECS)*, 2010, Wuhan, pp.1-4.
- [32] A. S. Ogrenci, Complex Event Post Processing for Traffic Accidents, *CINTI 2012, 13th IEEE International Symposium on Computational Intelligence and Informatics*, 20–22 November, 2012, Budapest, pp.341-345.
- [33] S. Li, D. Zhao, Prediction of Road Traffic Accidents Loss Using Improved Wavelet Neural Network, *Proceedings of IEEE TENCON'02*, 2002, pp. 1526-1529.
- [34] L. Yuejing, Z. Xing-lin, Z. Haixia, L. Ming, L. Jie, Research on Accident Prediction of Intersection and Identification Method of Prominent Accident Form Based on Back Propagation Neural Network, *International Conference on Computer Application and System Modeling (ICCASM 2010)*, 2010, pp. 434-438.
- [35] F. Huilin, Z. Yucai, The Traffic Accident Prediction Based on Neural Network, *Second International Conference on Digital Manufacturing & Automation*, 2011, pp.1349-1350.

- [36] P. Liu, A Neural Network Approach on Analyzing and Reducing Signalized Intersection Crashes, Proceedings of the Third International Conference on Natural Computation, ICNC '07, pp. 723-729.
- [37] W. Huiying, L. Jun, C. Xiaolong, G. Xiaohui, Real-time Highway Accident Prediction Based on Grey Relation Entropy Analysis and Probabilistic Neural Network, International Conference on Electric Technology and Civil Engineering (ICETCE), 2011, Lishan, pp.1420 – 1423.
- [38] A. Mahmoudabadi, Comparison of weighted and simple linear regression and artificial neural network models in freeway accidents prediction (Case study: Qom & Qazvin Freeways in Iran), Second International Conference on Computer and Network Technology, 2010, pp.392-396.
- [39] L.-Y. Chang, Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network, Safety Science, 43 (2005), pp.541–557.
- [40] L. Yuejing, S. Tang, H. Zhao, S. Li, Real-time Highway Accident Prediction based on Support Vector Machines, Chinese Control and Decision Conference (CCDC 2009), 2009, pp.4403-4407.
- [41] A. Theofilatos, G. Yannis, C. Antoniou, A. Chaziris & D. Sermpis (2017). Time series and support vector machines to predict powered-two-wheeler accident risk and accident type propensity: A combined approach. Journal of Transportation Safety & Security, 10(5), pp. 471–490. doi:10.1080/19439962.2017.1301611.
- [42] P. Sun, G. Guo & R. Yu (2017). Traffic crash prediction based on incremental learning algorithm. 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). doi:10.1109/icbda.2017.8078803.
- [43] M. Fayaz, I. Ullah & D.-H. Kim (2018). Underground Risk Index Assessment and Prediction Using a Simplified Hierarchical Fuzzy Logic Model and Kalman Filter. Processes, 6(8), 103. doi:10.3390/pr6080103.
- [44] A. M. Wahaballa, A. Diab, M. Gaber, A. M. Othman, Sensitivity of Traffic Accidents Mitigation Policies Based on Fuzzy Modeling: A Case Study, Proc. IEEE 20th Int. Conf. on Intelligent Transportation Systems (ITSC), pp 45–50, 2017.
- [45] M. Hosseinpour, A. S. Yahaya, S. M. Ghadiri, J. Prasetijo, Application of adaptive neuro-fuzzy inference system for road accident prediction, KSCE J Civ Eng, 17(7):1761–1772, 2013. <https://doi.org/10.1007/s12205-013-0036-3>.

- [46] R. Wang, Y. Chen, T. Li, P. Li, J. Sun (2013) Classification of road safety based on fuzzy clustering. Proc IEEE 10th Int Conf Fuzzy Syst Knowl Discov (FSKD), pp.354–358, 2009.
- [47] B. L. Xie, Z. Hu, and H. W. Ma, Fuzzy-logic-based traffic incident detection algorithm for freeway, in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2008, pp. 1254–1259.
- [48] M. Driss, T. Saint-Gerand, A. Bensaid, K. Benabdeli, M. A. Hamadouche, A fuzzy logic model for identifying spatial degrees of exposure to the risk of road accidents, *Int. Conf. Advanced Logistics and Transport (ICALT)*, 2013, Sousse, pp. 69 – 74.
- [49] L. Yong, Z. Shibo, The Fuzzy Regression Prediction of the City Road Traffic Accident, *International Conference on Industrial Mechatronics and Automation*, 2009, pp.121-124.
- [50] C. Bin, L. Yong, The Road Safety Prediction Model Based on the Fuzzy Linear Regression, *International Conference on Computational Intelligence and Natural Computing*, 2009, pp.19-21.
- [51] L. Lin, K. Yuan, S. Ren, Analysis of Urban Freeway Traffic Flow Characteristics based on Frequent Pattern Tree, *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014, pp.1719 – 1725.
- [52] A.-Z. Li and X.-H. Song, Traffic Accident Characteristics Analysis Based on Fuzzy Clustering, *IEEE Symposium on Electrical & Electronics Engineering (EESYM)*, 2012, pp.468-470.
- [53] J. Weng, J.Z. Zhu, X. Yan, Z. Liu, Investigation of work zone crash casualty patterns using association rules, *Accid Anal Prev*, pp.92:92–43, 2016.
- [54] Z. Gao, R. Pan, X. Wang, R. Yu, Research on automated modeling algorithm using association rules for traffic accidents, *Proc IEEE Int Conf Big DataSmart Comput*, pp.127–132, 2018.
- [55] S. Äyrämö, P. Pirtala, J. Kauttonen, K. Naveed, T. Karkkainen, Mining road traffic accidents, Technical Report, Reports of the Department of Mathematical Information Technology Series C. Software and Computational Eng. No.C. 2/2009, 2009, [http://users.jyu.fi/~samiayr/pdf/mining\\_road\\_traffic\\_accidents.pdf](http://users.jyu.fi/~samiayr/pdf/mining_road_traffic_accidents.pdf)
- [56] S. Kumar, D. Toshniwal, Analysing Road Accident Data Using Association Rule Mining, *Proc. of Int. Conf. on Computing Communication and Security*, pp 1–6, 2015.

- [57] V. A. Olutayo, A. A. Eludire, Traffic Accident Analysis Using Decision Trees and Neural Networks, *Int. J. Information Technology and Computer Science*, 2014, 2, pp.22-28.
- [58] T. Tambouratzis, D. Souliou, M. Chalikias and A. Gregoriades, Combining Probabilistic Neural Networks and Decision Trees for Maximally Accurate and Efficient Accident Prediction, The 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, pp. 2161-4393.
- [59] Y. Lv, S. Tang, H. Zhao, Real-time Highway Traffic Accident Prediction Based on the  $k$ -Nearest Neighbor Method, *International Conference on Measuring Technology and Mechatronics Automation*, 2009, pp.547-550.
- [60] L.-Y. Chang, and H.-W. Wang, Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques. *Accident Analysis and Prevention*, Vol. 38, No. 5, 2006, pp. 1019–1027.
- [61] T. Sayed, W. Abdelwahab, Comparison of Fuzzy and Neural Classifiers for Road Accidents Analysis, *Journal of Computing in Civil Engineering*, Vol. 12, No. 1, 1998, pp. 42–47.
- [62] H. T. Abdelwahab, M. A. Abdel-Aty. Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas, In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1784, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 115–125.
- [63] M. A. Abdel-Aty, and H. T. Abdelwahab. Predicting Injury Severity Levels in Traffic Crashes: A Modeling Comparison, *Journal of Transportation Engineering*, Vol. 130, No. 2, 2004, pp. 204–210.
- [64] M. Chong, A. Abraham, and M. Paprzycki. Traffic Accident Analysis Using Machine Learning Paradigms. *Informatica*, Vol. 29, No. 1, 2005, pp. 89–98.
- [65] B. Depaire, G. Wets, and K. Vanhoof, Traffic Accident Segmentation by Means of Latent Class Clustering. *Accident Analysis and Prevention*, Vol. 40, 2008, pp. 1257–1266.
- [66] J. L. McClelland, D. E. Rumelhart, *Explorations in Parallel Distributed Processing*, Cambridge, Massachusetts, USA: A Bradford Book, 1988.
- [67] B. Kosko, *Neural Networks and Fuzzy Systems*, Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [68] J. Zurada, *Introduction to Artificial Neural Systems*, West Publishing Company, St. Paul, MN, 1992.



- [69] B. Widrow and M. Lehr, 30 Years of Adaptive Neural Networks: Perceptron, Madaline and Backpropagation, *Proc. of the IEEE*, Vol. 78, No.9, pp. 1415-1442, 1990.
- [70] T. Kohonen, *The self-organizing map*, *Proc. IEEE*, vol. 78, pp. 1464-1480, Sept. 1990.
- [71] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*, Plenum, New York, 1981.
- [72] A. Baraldi and P. Blonda, A survey of fuzzy clustering algorithms for pattern recognition-part II, *IEEE Transactions on Systems, Man, and Cybernetics B*, -29, no. 6, pp.786-801, 1999.0.
- [73] J. S. R. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice Hall, 1997.
- [74] D. Gustafson, W. Kessel, Fuzzy clustering with a fuzzy covariance matrix, in: *Proc. IEEE CDC*, San Diego, CA, USA, pp. 761-766, 1979.
- [75] A. F. Kohn, Autocorrelation and Cross-Correlation Methods, in: *Encyclopedia of Biomedical Engineering*, Ed. Metin Akay, Hoboken, John Wiley & Sons, pp. 260-283, 2006.
- [76] J. Kennedy and R. C. Eberhart. Particle swarm optimization, *Proceedings of the 1995 IEEE International Conference on Neural Networks*, pp. 1942–1948, Piscataway, New Jersey, 1995. doi:<http://dx.doi.org/10.1109/ICNN.1995.488968>
- [77] M. Reyes-Sierra and C. A. Coello, Multi-objective particle swarm optimizers: A survey of the state-of-the-art, *International Journal of Computational Intelligence Research (IJCIR)*, 2:287–308, 2006.
- [78] J. Kennedy and R. C. Eberhart, *Swarm Intelligence*, Morgan Kaufmann Publishers, San Francisco, California, 2001.
- [79] T. A. Kumbhare and S. V. Chobe, An Overview of Association Rule Mining Algorithms, *International Journal of Computer Science and Information Technologies*, 5 (1), 927-930, 2014.
- [80] S. Brin, R. Motwani, J. D. Ullman, S. Tsur, Dynamic itemset counting and implication rules for market basket data, *Proceedings ACM SIGMOD International Conference on Management of Data*, Tucson, Arizona, USA, pp.255–264, 1997.

- [81] T. Brijs, G. Swinnen, K. Vanhoof, G. Wets, Building an association rules framework to improve product assortment decisions, *Data Mining and Knowledge Discovery*, 8(1), pp.7–23, 2004.
- [82] J. M. Luna, M. Ondra, H. M. Fardoun, S. Ventura, Optimization of quality measures in association rule mining: an empirical study, *International Journal of Computational Intelligence Systems*, Vol. 12, pp.59-78, 2018.
- [83] R. Agrawal, R. Srikant, Fast Algorithms for Mining Association Rules in Large Databases, *Proceedings of the 20th International Conference on Very Large Data Bases*, pp. 487-499, 1994.
- [84] Q. Zhao, S. S. Bhowmick, Association Rule Mining: A Survey, *Technical Report, CAIS, Nanyang Technological University, Singapore*, 2003.
- [85] R. Agrawal, T. Imielinski, A. N. Swami, Mining Association Rules between Sets of Items in Large Databases, *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Washington DC, 25-28 May 1993, Vol. 22(2) of SIGMOD Records, 207-216, 1993.
- [86] J. Han, P. Jian, and Y. Yiwen, M. Runying. Mining frequent patterns without candidate generation: A Frequent Pattern Tree Approach, *Data Mining and Knowledge Discovery*, 8, pp.53–87, 2004.
- [87] L. A. Zadeh, Possibility Theory and Soft Data Analysis, Cobb, L. and R.M. Thrall, eds., *Math. Frontiers of the Social and Policy Sciences*, pp. 69-129, Boulder, Colo.: Westview Press, 1981.
- [88] L. A. Zadeh, Fuzzy logic, neural networks, and soft computing, *Communications of the ACM*, vol. 37, no.3, pp.77-84, March 1994.
- [89] G. J. Klir, B. Yuan, Fuzzy sets and fuzzy logic: theory and applications, Prentice-Hall, Inc., Upper Saddle River, NJ, 1994.
- [90] D. Kukulj, Sistemi zasnovani na računarskoj inteligenciji, monografija, *FTN izdavaštvo*, 2007.
- [91] M. Sugeno and T. Yasukawa, A fuzzy-logic-based approach to qualitative modeling, *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 7-31, Feb. 1993.
- [92] L. X. Wang and J. M. Mendel, Generating fuzzy rules by learning from examples, *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 6, pp. 1414-1427, 1992.
- [93] D. Kukulj, E. Levi, Identification of Complex Systems Based on Neural and Takagi-Sugeno Fuzzy Model, *IEEE IEEE Trans. Syst., Man, and Cybernetics - part B*, Vol. 34, No. 1, pp.272-282, February 2004.

- [94] J. Yen, L. Wang and C. W. Gillespie, Improving the interpretability of TSK fuzzy models by combining global learning and local learning, *IEEE Trans. Fuzzy Syst.*, vol. 6, no. 4, pp. 531-537, Nov. 1998.
- [95] D. Kukolj, B. Atlagić, M. Petrov, Unlabeled data clustering using a re-organizing neural network, *Cybernetics and Systems, An Int. Journal*, Vol. 37, No. 7, pp. 779-790, 2006. doi:10.1080/01969720600887152.
- [96] G. H Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1989.
- [97] D. P. Pancho, J. M. Alonso, O. Cordon, A. Quirin, L. Magdalena, FINGRAMS: Visual Representations of Fuzzy Rule-Based Inference for Expert Analysis of Comprehensibility, *IEEE Trans. Fuzzy Syst*, 21(6):1133-1149, 2013.
- [98] D. R. Ferreira, *A Primer on Process Mining*, Springer, 2017, doi:10.1007/978-3-319-56427-2.
- [99] T. Takagi, M. Sugeno, Fuzzy Identification of Systems and Its Application to Modelling and Control. *IEEE Trans. Syst. Man Cybern*, vol.15, pp.116-132, 1985.
- [100] A. Bulajić, D.Jovanović, B. Matović, S.D. Bačkalić, Identification of high-density locations with homogenous attributes of pedestrian accident in the urban area of Novi Sad. XII Int. Symp, Road Accidents Preventions, pp.89–98, 2014.
- [101] L. van der Maaten and G. Hinton, Visualizing data using t-SNE, *Journal of Machine Learning Research*, vol. 9, pp. 2579-2605, Nov. 2008.
- [102] S. Nemet, D. Kukolj, G. Ostojić, S. Stankovski, D. Jovanović, Aggregation framework for TSK fuzzy and association rules: interpretability improvement on a traffic accidents case, *Appl Intell*, 2019, <https://doi.org/10.1007/s10489-019-01485-6>.
- [103] J. M. Alonso, L. Magdalena, HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers, *Soft Comput*, 15(10), pp.1959–1980, 2011, DOI 10.1007/s00500-010-0628-5.
- [104] D. D. Nauck, Measuring interpretability in rule-based classification systems, *Proc. of 12th IEEE INT CONF FUZZY'03*, pp. 196–201, 2003.
- [105] D. P. Pancho, J.M. Alonso, L. Magdalena, Quest for interpretability-accuracy trade-off supported by Fingrams into the fuzzy modeling tool GUAJE, *Int. Journal of Computational Intelligence Systems*, 6 (1), pp. 46-60, 2013, DOI:10.1080/18756891.2013.818189.

[106] J. M. Alonso, L. Magdalena, Generating understandable and accurate fuzzy rule-based systems in a java environment, LECT NOTES ARTIF INT - 9th Int. Workshop on Fuzzy Logic and Applications, LNAI6857 pp. 212-219, 2011, DOI:10.1007/978-3-642-23713-3\_27.