



UNIVERZITET U NOVOM SADU

FAKULTET TEHNIČKIH NAUKA



KLJUČNI FAKTORI PRIMENE AUTOMATSKOG OTKRIVANJA POSLOVNIH PROCESA U INDUSTRIJSKIM SISTEMIMA

DOKTORSKA DISERTACIJA

Mentor:

prof. dr Darko Stefanović
prof. dr Branislav Stevanov

Kandidat:

Dušanka Dakić

Novi Sad, 2023. godine

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА¹

Врста рада:	Докторска дисертација
Име и презиме аутора:	Душанка Дакић
Ментор (титула, име, презиме, звање, институција)	др Дарко Стефановић, редовни професор, Факултет техничких наука, Универзитет у Новом Саду др Бранислав Стеванов, ванредни професор, Факултет техничких наука, Универзитет у Новом Саду
Наслов рада:	Кључни фактори примене аутоматског откривања пословних процеса у индустријским системима
Језик публикације (писмо):	Српски (латиница)
Физички опис рада:	Унети број: Страница 158 Поглавља 7 Референци 318 Табела 65 Слика 20 Графикона 0 Прилога 2
Научна област:	Индустријско инжењерство и инжењерски менаџмент
Ужа научна област (научна дисциплина):	Информационо–комуникациони системи
Кључне речи / предметна одредница:	Аутоматско откривање пословних процеса, дневник извршења догађаја, проблеми квалитета података, откривање пословних процеса
Резиме на језику рада:	У оквиру докторске дисертације истражује се повезаност кључних фактора аутоматског откривања пословних процеса и њихов утицај на ток извршавања пројеката анализе пословних процеса. Такође, испитује се и повезаност између проблема квалитета података и техника чишћења дневника извршења догађаја. На основу резултата испитаних хипотеза истраживачких модела, утврђени су кључни фактори примене у индустријским системима и дефинисане су смернице за планирање и извођење аутоматског откривања пословних процеса.
Датум прихватања теме од стране надлежног већа:	25.03.2021.
Датум одбране: (Попуњава одговарајућа служба)	
Чланови комисије: (титула, име, презиме, звање, институција)	Председник: др Теодора Вучковић, доцент, Факултет техничких наука, Универзитет у Новом Саду Члан: др Милан Мирковић, ванредни професор, Факултет техничких наука, Универзитет у Новом Саду Члан: др Душан Савић, ванредни професор, Факултет организационих наука, Универзитет у Београду Ментор: др Дарко Стефановић, редовни професор, Факултет техничких наука, Универзитет у Новом Саду Ментор: Др Бранислав Стеванов, ванредни професор, Факултет техничких наука, Универзитет у Новом Саду
Напомена:	

¹Аутор докторске дисертације потписао је и приложио следеће Обрасце:

5б – Изјава о ауторству;

5в – Изјава о истоветности штампане и електронске верзије и о личним подацима;

5г – Изјава о коришћењу.

Ове Изјаве се чувају на факултету у штампаном и електронском облику и не кориче се са тезом.

KEY WORD DOCUMENTATION²

Document type:	Doctoral dissertation
Author:	Dusanka Dakic
Supervisor (title, first name, last name, position, institution)	dr Darko Stefanovic, full professor, Faculty of Technical Sciences, University of Novi Sad dr Branislav Stevanov, associate professor, Faculty of Technical Sciences, University of Novi Sad
Thesis title:	The key factors of process mining implementation in industrial systems
Language of text (script):	Serbian language (latin script)
Physical description:	Number of: Pages 158 Chapters 7 References 318 Tables 65 Illustrations 20 Graphs 0 Appendices 2
Scientific field:	Industrial Engineering and Engineering Management
Scientific subfield (scientific discipline):	Information and Communication Systems
Subject, Keywords:	Process mining, event log, data quality issues, process discovery
Abstract in English language:	This doctoral dissertation studies the relationship between the key factors of process mining projects and their influence on the project execution process. Furthermore, the relationship between event log data quality issues and preprocessing techniques is examined. Based on the results of research models' hypothesis testing, the key factors of process mining projects in industrial systems are defined and the guidelines for planning and executing process mining projects, as well as guidelines for event log preprocessing are developed.
Accepted on Scientific Board on:	25.03.2021.
Defended: (Filled by the faculty service)	
Thesis Defend Board: (title, first name, last name, position, institution)	President: dr Teodora Vučković, assistant professor, Faculty of Technical Sciences, University of Novi Sad Member: dr Milan Mirkovic, associate professor, Faculty of Technical Sciences, University of Novi Sad Member: dr Dušan Savić, associate professor, Faculty of Organizational Sciences, University of Belgrade Supervisor: dr Darko Stefanović, full professor, Faculty of Technical Sciences, University of Novi Sad Supervisor: dr Branislav Stevanov, associate professor, Faculty of Technical Sciences, University of Novi Sad
Note:	

² The author of doctoral dissertation has signed the following Statements:

56 – Statement on the authority,

5B – Statement that the printed and e-version of doctoral dissertation are identical and about personal data,

5r – Statement on copyright licenses.

The paper and e-versions of Statements are held at the faculty and are not included into the printed thesis.

Zahvalnost

Želim da se zahvalim mentorima, prof. dr Darku Stefanoviću i prof. dr Branislavu Stevanovu na iskrenoj pomoći i podršci u naučnom radu i prilikom pisanja disertacije.

Takođe, zahvaljujem se kolegama sa Departmana za industrijsko inženjerstvo i menadžment, koji su svojim sugestijama pružili podršku prilikom razvoja idejnog koncepta i sprovođenja istraživanja.

Posebnu zahvalnost dugujem roditeljima.

Sadržaj

Rezime.....	iii
Summary.....	iv
Lista slika.....	v
Lista tabela.....	vi
Lista skraćenica.....	viii
1. Uvodna razmatranja.....	1
1.1. Motivacija, definisanje i opis predmeta istraživanja	2
1.2. Cilj istraživanja, istraživačka pitanja i hipoteze	4
1.3. Tok istraživanja doktorske disertacije	5
1.4. Rezultati istraživanja i naučni doprinos disertacije	7
1.5. Struktura disertacije.....	8
2. Teorijske osnove	10
2.1. Osnovni koncepti automatskog otkrivanja poslovnih procesa	10
2.1.1. Definicija automatskog otkrivanja poslovnih procesa	10
2.1.2. Dnevnik izvršenja događaja	11
2.1.3. Tipovi i perspektive automatskog otkrivanja poslovnih procesa.....	12
2.2. Pregled stanja u oblasti.....	15
2.2.1. Pregled primene automatskog otkrivanja poslovnih procesa u industrijskim sistemima.....	15
2.2.2. Pregled primene automatskog otkrivanja poslovnih procesa u proizvodnim industrijskim sistemima	19
2.2.3. Pregled problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja.....	24
2.2.3.1. Sinteza problema kvaliteta podataka dnevnika izvršenja događaja	24
2.2.3.2. Sinteza tehnika čišćenja dnevnika izvršenja događaja.....	26
2.3. Rezultati prethodnih istraživanja.....	30
2.3.1. Postojeće metodologije i uputstva	30
2.3.2. Prethodni pregledi literature	31
2.4. Istraživački modeli i hipoteze.....	33
3. Metodološki aspekti istraživanja	36
3.1. Metodologija prikupljanja i obrade podataka istraživačkog modela I.....	36
3.1.1. Protokol pregleda literature	37
3.1.2. Izvođenje pregleda literature	39
3.1.3. Bibliografski podaci primarnih studija	42
3.1.4. Primijenjene metode statističke obrade podataka	46
3.2. Metodologija prikupljanja i obrade podataka istraživačkog modela II	48
3.2.1. Razvoj i distribucija mernog instrumenta.....	48
3.2.2. Demografske karakteristike ispitanika	50
3.2.3. Primijenjene metode statističke obrade podataka	56
4. Rezultati istraživanja.....	59
4.1. Rezultati istraživačkog modela I	59
4.1.1. Deskriptivna statistika varijabli istraživačkog modela I.....	59
4.1.1.1. Industrijski sistem, cilj analize i poslovni procesi	59
4.1.1.2. Tipovi automatskog otkrivanja poslovnih procesa.....	63

4.1.1.3. Algoritmi, tehnike i softverski alati	64
4.1.1.4. Pridružene tehnologije	69
4.1.2. Rezultati ispitivanja hipoteza istraživačkog modela I	70
4.1.2.1. Rezultat ispitivanja hipoteze H1: Povezanost vrste industrije sa vrstom poslovnog procesa	70
4.1.2.1. Rezultat ispitivanja hipoteze H2: Povezanost vrste industrije sa ciljem analize automatskog otkrivanja poslovnih procesa	71
4.1.2.2. Rezultat ispitivanja hipoteze H3: Povezanost cilja analize automatskog otkrivanja poslovnih procesa sa vrstom poslovnog procesa	71
4.1.2.3. Rezultat ispitivanja hipoteze H4: Povezanost cilja analize automatskog otkrivanja poslovnih procesa sa tipom	71
4.1.2.4. Rezultat ispitivanja hipoteze H5: Povezanost cilja analize automatskog otkrivanja poslovnih procesa sa primenjenim algoritmima i tehnikama	72
4.1.3. Tabela kontingencije odnosa između vrste proizvodnje, cilja analize i tipa automatskog otkrivanja poslovnih procesa	74
4.2. Rezultati istraživačkog modela II	75
4.2.1. Percipirani značaj i učestalost susretanja problema kvaliteta dnevnika izvršenja događaja	76
4.2.2. Percipirani značaj i učestalost primene tehnika čišćenja dnevnika izvršenja događaja	78
4.2.3. IPA analiza	79
4.2.3.1. IPA analiza percipiranog značaja i učestalosti susretanja problema kvaliteta podataka dnevnika izvršenja događaja	79
4.2.3.2. IPA analiza percipiranog značaja i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja	82
4.2.4. Analiza varijansi	84
4.2.4.1. Razlike u značaju koji ispitanici različitih uloga pridaju problemima kvaliteta podataka dnevnika izvršenja događaja i tehnikama čišćenja dnevnika izvršenja događaja	84
4.2.4.2. Razlike u pogledu učestalosti susreta ispitanika različitih uloga sa problemima kvaliteta podataka i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja	85
4.2.4.3. Razlike u značaju koji ispitanici različitog nivoa iskustva pridaju problemima kvaliteta podataka dnevnika izvršenja događaja i tehnikama čišćenja dnevnika izvršenja događaja	87
4.2.5. Rezultat ispitivanja hipoteze istraživačkog modela II	87
4.2.5.1. Rezultat Pirsonovog testa korelacije između učestalosti susreta problema kvaliteta podataka i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja	88
4.2.5.2. Rezultat Hi-kvadrat testa povezanosti problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja	89
5. Diskusija rezultata istraživanja	94
5.1. Diskusija rezultata istraživačkog modela I	94
5.1.1. Diskusija procedure prikupljanja podataka i deskriptivne statistika varijabli istraživačkog modela I	94
5.1.2. Diskusija rezultata ispitivanja hipoteza istraživačkog modela I	97
5.1.3. Smernice za planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa u zavisnosti od vrste industrije i cilja analize	99
5.2. Diskusija rezultata istraživačkog modela II	102
5.2.1. Diskusija razvoja mernog instrumenta i demografije ispitanika	103
5.2.2. Diskusija analize značaja i učestalosti problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja	104
5.2.3. Diskusija rezultata ispitivanja hipoteze istraživačkog modela II	105
5.2.4. Smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu	106
6. Zaključna razmatranja i pravci daljih istraživanja	106
7. Literatura	109
Prilog A Spisak primarnih studija	130
Prilog B Upitnik	145

Rezime

Automatsko otkrivanje poslovnih procesa je disciplina koja ima za cilj da analizira podatke izvučene iz informacionih sistema koji podržavaju izvršavanje poslovnih procesa, kako bi se definisali modeli procesa onako kako se izvršavaju u realnom sistemu i uočili problemi u performansama procesa. Podaci o izvršavanju procesa se transformišu u dnevnik izvršenja događaja, gde nivo kvaliteta podataka direktno utiče na pouzdanost, validnost i primenljivost stečenih uvida o izvršavanju procesa. Iako postoji veliki broj softverskih alata za automatsko otkrivanje poslovnih procesa, kao i pristupa, algoritama i tehnika, ne postoji velik broj instrukcija i uputstava na temu odabira pogodnog pristupa prilikom izvođenja projekata automatskog otkrivanja poslovnih procesa. Takođe, postojeće metodologije automatskog otkrivanja poslovnih procesa se ne bave problemima odabira odgovarajuće tehnike čišćenja dnevnika izvršenja događaja u odnosi na odgovarajući problem kvaliteta podataka. Postojeća literatura sadrži nekoliko metodologija za izvođenje automatskog otkrivanja poslovnih procesa sa nedovoljno detaljnim, generičkim koracima, taksonomiju tehnika čišćenja dnevnika izvršenja događaja i radove koji izveštavaju o rešenjima za konkretan problem kvaliteta podataka, bez ispitivanja veze između problema kvaliteta podataka i mogućih rešenja.

Ovo istraživanje teži da otkrije koji su ključni faktori automatskog otkrivanja poslovnih procesa koji utiču na planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa. Takođe, istraživanje ispituje kako istraživači i praktikanti rešavaju probleme kvaliteta podataka i postojanje veze između problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja. Stoga, definisana su dva konkurentna istraživačka modela.

Kako bi se otkrilo koji su ključni faktori automatskog otkrivanja poslovnih procesa koji utiču na planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa, sproveden je sistematski pregled literature automatskog otkrivanja poslovnih procesa primena u različitim industrijskim sistemima. Rezultati su pokazali da su ključni faktori vrsta industrije, cilj analize, vrsta poslovnog procesa, tip automatskog otkrivanja poslovnih procesa i odgovarajući algoritmi i tehnike. Na osnovu potvrđene, statistički značajne veze između vrste industrije i vrste poslovnog procesa, vrste industrije i cilja analize, cilj analize i tipa automatskog otkrivanja poslovnih procesa, algoritama i tehnika, definisane su smernice za planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa u zavisnosti od vrste industrije i cilja analize.

Dalje, empirijsko istraživanje je sprovedeno među prominentnim istraživačima i praktikantima automatskog otkrivanja poslovnih procesa, kako bi se prikupile informacije o percipiranom značaju i učestalosti susreta i primene problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja. Rezultati pokazuju koji su najznačajniji u najučestaliji problemi kvaliteta podataka i tehnike čišćenja dnevnika izvršenja događaja, razliku u njihovom značaju i učestalosti, i potvrđuju statistički značajnu vezu između problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja. Na osnovu predstavljenih rezultata, definisane su smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu.

Rezultati doktorske disertacije pružaju podršku praktikantima i istraživačima automatskog otkrivanja poslovnih procesa prilikom odabira odgovarajućeg cilja analize i pristupa izvođenja analize poslovnih procesa u industrijskim sistemima.

Summary

Process mining is a discipline that analyses real event data extracted from information systems that support a business process to construct as-is process models and detect performance issues. Process event data are transformed into event logs, where the level of data quality directly impacts the reliability, validity, and usefulness of the derived process insights. Due to the novelty of the field, there are numerous process mining tools, types, algorithms and techniques that can be applied, with a low number of specific instructions and guidelines on how to select the suitable approach when conducting process mining projects. Furthermore, current process mining methodologies do not tackle the issue of the selection of suitable preprocessing techniques regarding a specific data quality issue. The literature offers several methodologies with generic steps, a taxonomy of preprocessing techniques, and papers reporting on solutions for data quality issues in particular scenarios without exploring the relationship between the data quality issues and solutions.

This research aims to discover what are the key factors of process mining projects that influence the planning and execution steps of process mining. Furthermore, the research questions how process mining researchers and practitioners solve certain data quality issues in practice and investigates the nature of the relationship between data quality issues and pre-processing techniques. Therefore, two concurrent research models were developed.

To discover the key factors of process mining projects that influence the planning and execution steps of process mining, a systematic literature review of process mining applications in was examined. The results conclude that the key factors are industry type, analysis goal, business process type, process mining type, and algorithms and techniques. Based on confirming a statistically significant relationship between industry type and business process type, industry type and analysis goal, analysis goal and process mining types and algorithms, a guideline for the planning and execution of process mining projects is developed.

Furthermore, a study was obtained among prominent process mining researchers and practitioners, gathering information regarding the perceived importance and frequency of data quality issues and solutions and the participants' recommendations on preprocessing technique selection. The results reveal the most important and frequent data quality issues, preprocessing techniques, the gap between their perceived frequency and importance, and confirm a statistically significant relationship between data quality issues and preprocessing techniques. Consequently, the overview of how researchers and practitioners solve data quality issues is presented, allowing the development of recommendations.

The results of the doctoral dissertation are aiding process mining practitioners and researchers to properly select suitable analysis goals and select the suitable approach while performing business process analysis in industrial systems.

Lista slika

Slika 1. Osnovni koncepti automatskog otkrivanja poslovnih procesa [2]	2
Slika 2. Tok istraživanja doktorske disertacije	6
Slika 3. Životni ciklus projekta automatskog otkrivanja poslovnih procesa [1]	11
Slika 4. Petrijeva mreža modela procesa dobijena na osnovu primene α -algoritma na dnevnik D1 [3]	14
Slika 5. Udeo primene tipova automatskog otkrivanja poslovnih procesa u industriji	16
Slika 6. Industrije u posmatranim primarnim studijama u kojima je primenjeno automatsko otkrivanje poslovnih procesa i njihov udeo	17
Slika 7. Ciljevi primene automatskog otkrivanja poslovnih procesa u proizvodnoj industriji	21
Slika 8. Kombinacije primene tipova automatskog otkrivanja poslovnih procesa u proizvodnoj industriji	21
Slika 9. Distribucija poslovnih procesa proizvodne industrije analiziranih u posmatranim studijama slučaja	22
Slika 10. Istraživački model 1	33
Slika 11. Istraživački model 2	34
Slika 12. <i>PRISMA</i> tok izvršavanja identifikacije primarnih studija	41
Slika 13. Histogram učestalosti objavljivanja primarnih studija po godinama	42
Slika 14. Broj objavljenih vrsta radova prema godinama	43
Slika 15. Mreža saradnje između autora primarnih studija	45
Slika 16. Mreža povezanosti ključnih reči primarnih studija	46
Slika 17. Zastupljenost ispitanika uzorka prema zanimanju	52
Slika 18. Mapa zastupljenosti ispitanika po državama	53
Slika 19. Dijagram rasprostranjenosti problema kvaliteta podataka dnevnika izvršenja događaja na osnovu percipiranog značaja i učestalosti susretanja	82
Slika 20. Dijagram rasprostranjenosti tehnika čišćenja podataka dnevnika izvršenja događaja na osnovu percipiranog značaja i učestalosti primene	84

Lista tabela

Tabela 1. Primer dnevnika izvršenja događaja [31]	11
Tabela 2. Matrica sledljivosti aktivnosti dnevnika <i>DI</i> – „Otisak”	13
Tabela 3. Manifestacija problema kvaliteta podataka i obrazaca imperfekcije.....	24
Tabela 4. Grupisane tehnike čišćenja dnevnika izvršenja događaja prema literaturi	29
Tabela 5. Prethodni pregledi literature iz oblasti automatskog otkrivanja poslovnih procesa	32
Tabela 6. Stavke koje određuju dimenzije ključnih činilaca istraživanja u okviru IM2	35
Tabela 7. Strategija prikupljanja podataka	38
Tabela 8. Učestalost i procenat primarnih studija po godinama.....	42
Tabela 9. Učestalost i procenat primarnih studija prema vrsti rada	43
Tabela 10. Najučestaliji časopisi u kojima su objavljene primarne studije.....	43
Tabela 11. Naučestaliji izvodi konferencija u kojima su objavljene primarne studije	44
Tabela 12. Grupisani autori primarnih studija sa najviše radova	45
Tabela 13. Zastupljenost ispitanika uzorka prema iskustvu u obradi podataka	50
Tabela 14. Zastupljenost ispitanika uzorka prema iskustvu u oblasti automatskog otkrivanja poslovnih procesa.....	51
Tabela 15. Tabela kontingencije odnosa uloge i vremenskog perioda izučavanja/primene automatskog otkrivanja poslovnih procesa.....	51
Tabela 16. Zastupljenost ispitanika uzorka prema kontinentima na kojima su zaposleni	52
Tabela 17. Zastupljenost ispitanika uzorka prema softverskom alatu za primenu automatskog otkrivanja poslovnih procesa.....	53
Tabela 18. Zastupljenost ispitanika uzorka prema softverskom alatu za obradu dnevnika izvršenja događaja	54
Tabela 19. Tabela kontingencije odnosa uloge i softverskih alata za primenu automatskog otkrivanja poslovnih procesa.....	54
Tabela 20. Tabela kontingencije odnosa uloge i softverskih alata za obradu podataka u kontekstu automatskog otkrivanja poslovnih procesa	55
Tabela 21. Učestalost i udeo različitih vrsta industrijskih sistema.....	59
Tabela 22. Učestalost i udeo različitih ciljeva analize.....	59
Tabela 23. Učestalost i udeo različitih poslovnih procesa.....	60
Tabela 24. Učestalost i udeo različitih tipova automatskog otkrivanja poslovnih procesa	63
Tabela 25. Učestalost i udeo različitih algoritama za otkrivanje poslovnih procesa.....	64
Tabela 26. Učestalost i udeo softverskih alata za otkrivanje poslovnih procesa.....	65
Tabela 27. Učestalost i udeo tehnika predviđanja u automatskom otkrivanju poslovnih procesa	66
Tabela 28. Učestalost i udeo tehnika za unapređenje poslovnih procesa	66
Tabela 29. Učestalost i udeo softverskih alata za unapređenje poslovnih procesa	67
Tabela 30. Učestalost i udeo tehnika za proveru usaglašenosti.....	68
Tabela 31. Učestalost i udeo softverskih alata za proveru usaglašenosti	68
Tabela 32. Učestalost i udeo softverskih alata za analizu društvenih mreža.....	69
Tabela 33. Učestalost i udeo pridruženih tehnologija automatskog otkrivanja poslovnih procesa.....	69
Tabela 34. Rezultat ispitivanja povezanosti vrsta industrije ↔ vrsta poslovnog procesa.....	70
Tabela 35. Rezultat ispitivanja povezanosti vrsta industrije ↔ cilj analize	71
Tabela 36. Rezultat ispitivanja povezanosti cilj analize ↔ vrsta poslovnog procesa	71
Tabela 37. Rezultat ispitivanja povezanosti cilj analize ↔ tip automatskog otkrivanja poslovnih procesa	72
Tabela 38. Rezultat ispitivanja povezanosti cilj analize ↔ algoritmi za otkrivanje poslovnih procesa	72
Tabela 39. Rezultat ispitivanja povezanosti cilj analize ↔ tehnike unapređenja poslovnih procesa	72
Tabela 40. Rezultat ispitivanja povezanosti cilj analize ↔ tehnike provere usaglašenosti procesa	73
Tabela 41. Rezultat ispitivanja povezanosti cilj analize ↔ tehnike analize društvenih mreža	73
Tabela 42. Rezultat ispitivanja povezanosti cilj analize ↔ tehnike predviđanja	73

Tabela 43. Tabela kontingencije odnosa između vrste proizvodnje, cilja analize i tipa automatskog otkrivanja poslovnih procesa.....	74
Tabela 44. Percipirani značaj problema kvaliteta podataka dnevnika izvršenja događaja.....	76
Tabela 45. Percipirana učestalost problema kvaliteta podataka dnevnika izvršenja događaja	77
Tabela 46. Percipirani značaj primene tehnika čišćenja dnevnika izvršenja događaja	78
Tabela 47. Percipirana učestalost primene tehnika čišćenja dnevnika izvršenja događaja	78
Tabela 48. Aritmetička srednja vrednost percipiranog značaja i učestalosti problema kvaliteta dnevnika izvršenja događaja	79
Tabela 49. Aritmetička srednja vrednost percipiranog značaja i učestalosti tehnika čišćenja dnevnika izvršenja događaja	83
Tabela 50. Razlike u značaju koji ispitanici različitih uloga pridaju problemima kvaliteta podataka dnevnika izvršenja događaja	85
Tabela 51. Razlike u značaju koji ispitanici različitih uloga pridaju tehnikama čišćenja dnevnika izvršenja događaja	85
Tabela 52. Razlike u pogledu učestalosti susreta ispitanika različitih uloga sa problemima kvaliteta podataka	86
Tabela 53. Razlike u pogledu učestalosti primene tehnika čišćenja dnevnika izvršenja događaja od strane ispitanika različitih uloga	86
Tabela 54. Razlike u značaju koji ispitanici različitog nivoa iskustva pridaju problemima kvaliteta podataka dnevnika izvršenja događaja	87
Tabela 55. Razlike u značaju koji ispitanici različitog nivoa iskustva pridaju tehnikama čišćenja dnevnika izvršenja događaja	87
Tabela 56. Rezultati Pirsonovog testa korelacije – Učestalost susreta problema kvaliteta podataka i učestalost primene tehnika čišćenja dnevnika izvršenja događaja	88
Tabela 57. Rezultat ispitivanja povezanosti problema kvaliteta podataka dnevnika izvršenja događaja i tehnika čišćenja dnevnika izvršenja događaja.....	89
Tabela 58. Vrednost Kramerovog koeficijenta veze između problema kvaliteta podataka dnevnika izvršenja događaja i tehnika čišćenja dnevnika izvršenja događaja.....	89
Tabela 59. Tabela kontingencije odnosa između problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja u procentima	90
Tabela 60. Poređenje zastupljenosti primene automatskog otkrivanja poslovnih procesa u različitim industrijama i vremenskim periodima.....	95
Tabela 61. Poređenje zastupljenosti postavljenih ciljeva analize automatskog otkrivanja poslovnih procesa u različitim vremenskim periodima	96
Tabela 62. Poređenje zastupljenosti primene tipova automatskog otkrivanja poslovnih procesa u različitim vremenskim periodima.....	97
Tabela 63. Smernice za planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa u zavisnosti od vrste industrije i cilja analize.....	99
Tabela 64. Smernice za odabir algoritma i tehnike različitih tipova automatskog otkrivanja poslovnih procesa.....	102
Tabela 65. Smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu	106

Lista skraćenica

Srpski jezik

AOPP – Automatsko otkrivanje poslovnih procesa

P – Problem

IP – Istraživačko pitanje

SPL – Sistematski pregled literature

OI – Obrazac imperfekcije

IM – Istraživački model

GPS – Globalni Pozicioni Sistem

Engleski jezik

PAIS – engl. *Process-aware Information Systems*

BPM – engl. *Business Process Management*

BPMN – engl. *Business Process Model and Notation*

IEEE – engl. *Institute of Electrical and Electronics Engineers*

DAAAM – engl. *Danube Adria Association For Automation & Manufacturing*

APMS – engl. *Advances in Production Management Systems*

IPA – engl. *Importance-Performance Analysis*

ANOVA – engl. *Analysis of Variance*

XES – engl. *eXtensible Event Stream*

W3C – engl. *World Wide Web Consortium*

XML – engl. *Extensible Markup Stream*

DFG – engl. *Directly-follows graph*

ERP – engl. *Enterprise Resource Planning*

MES – engl. *Manufacturing Execution System*

LSTM – engl. *Long Short-Term Memory*

SIER – engl. *Single Imputation by Event Relationship*

MIEC – engl. *Multiple Imputation by Event Chain*

CART – engl. *Classification and Regression Tree*

MST – engl. *Minimum-Spanning Tree*

NLP – engl. *Natural Language Processing*

ILP – engl. *Integer Linear Programming*

df – engl. *degree of freedom*

CEO – engl. *Chief Executive Officer*

SQL – engl. *Structured Query Language*

AWS – engl. *Amazon Web Services*

LSD – engl. *Least Significant Difference*

LTL – engl. *Linear Temporal Logic*

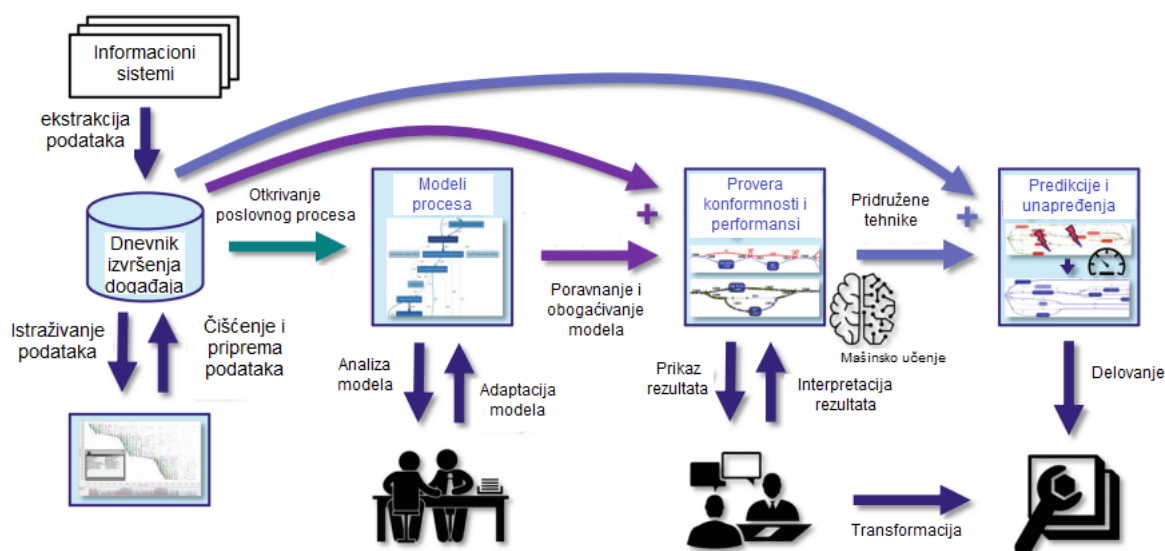
TDABC – engl. *Time-drive activity-based costing*

GERT – engl. *Graphical Evaluation and Review Technique*

1. UVODNA RAZMATRANJA

Trendovi u automatizaciji poslovnih procesa i upravljanju znanjem o interakcijama unutar procesa su postali ključ digitalne transformacije jedne organizacije. Poslovni procesi predstavljaju srž poslovanja organizacija još od prihvatanja procesnog pogleda na organizaciju. Poslovna organizacija se posmatra kao procesni sistem sačinjen od skupa poslovnih procesa, gde procesi ostvaruju međusobnu interakciju radi ostvarivanja zajedničkog cilja. S obzirom na značaj poslovnih procesa i današnje visoko automatizovano poslovno okruženje, postoje oblasti koje se bave analizom podataka koje generišu procesno-orijentisani informacioni sistemi (engl. *Process-Aware Information Systems* – PAIS). Procesno-orijentisani informacioni sistemi (ali i informacioni sistemi koji podržavaju izvođenje poslovnih procesa, ali nisu eksplicitno procesno orijentisani) generišu velike količine podataka o izvođenju poslovnih procesa koji se mogu analizirati radi unapređenja poslovanja. Pored discipline nauke o podacima (engl. *data science*), koja je interdisciplinarno polje i pruža alate za obradu tih podataka u informacije od vrednosti, a kojoj pripadaju oblasti statistika, istraživanje i eksploatacije podataka (engl. *data mining*), veštačka inteligencija (engl. *artificial intelligence*), mašinsko učenje (engl. *machine learning*) i poslovna inteligencija (engl. *business intelligence*), početkom 2000-tih godina se javlja ideja o zasebnoj disciplini koja bi mogla da spoji znanje nauke o podacima sa znanjem nauke o procesima (engl. *process science*) [1,2], sa nazivom automatsko otkrivanje poslovnih procesa – AOPP (engl. *process mining*). Nauka o procesima je širok pojam koji obuhvata znanje iz oblasti informaciono-komunikacionih tehnologija – IKT (engl. *Information and Communication Technology* – ICT) i upravljanja poslovnim procesima (engl. *Business Process Management* – BPM) kako bi se analizirali i unapredili operativni poslovni procesi korišćenjem informacionih tehnologija [3].

Automatsko otkrivanje poslovnih procesa pre svega omogućava organizacijama da otkriju način izvršavanja poslovnih procesa automatski generišući modele poslovnih procesa na osnovu podataka o izvršavanju aktivnosti koje čine neki poslovni proces. Slika 1 predstavlja osnovne koncepte automatskog otkrivanja poslovnih procesa i uobičajeni tok izvršavanja analize nekog poslovnog procesa primenom njegovih tehnika [2]. Podatke o izvršavanju aktivnosti je u određenoj meri kvaliteta moguće, uz veće ili manje poteškoće, ekstrahovati iz bilo kojeg informacionog sistema i transformisati u format dnevnika izvršenja događaja. Sam tok ekstrakcije, čišćenja i pripreme podataka zavisi od tipa informacionog sistema koji podržava izvršavanje poslovnog procesa, kao i od prirode i kvaliteta zabeleženih podataka. Postoji nekoliko tipova automatskog otkrivanja poslovnih procesa koji se mogu primeniti na pripremljen dnevnik izvršenja događaja. Prvi i najčešće primenjivani tip jeste otkrivanje poslovnog procesa (engl. *process discovery*), koje podrazumeva primenu algoritama na podatke dnevnika izvršenja događaja kako bi se izmodelovao tok izvršavanja aktivnosti unutar analiziranog procesa [1–3]. Rezultat otkrivanja poslovnog procesa jeste model procesa predstavljen kroz neku od podržanih notacija, poput Petrijevih mreža ili BPMN notacije za modelovanje poslovnih procesa (engl. *Business Process Model and Notation* – BPMN). Vremenom su se razvili i drugi tipovi automatskog otkrivanja poslovnih procesa, poput unapređenja poslovnog procesa (engl. *process enhancement*), koje proširuje model procesa sa informacijama o vremenu trajanja izvršavanja aktivnosti i vremenu čekanja između aktivnosti kako bi se pronašli problemi u izvršavanju procesa i omogućilo davanje predloga za njegovo unapređenje. Zatim, analiza performansi procesa (engl. *performance analysis*) i provera usaglašenosti (engl. *conformance checking*) između modela procesa i dnevnika izvršenja događaja, kako bi se utvrdio stepen kvaliteta dobijenog modela procesa. U poslednjih nekoliko godina razvio se pristup „gledanja unapred” (engl. *forward-looking*), koji primenjuje neke od pridruženih tehnika nauke o podacima, od kojih je najznačajnije mašinsko učenje. Mašinsko učenje je moguće primeniti na delove dnevnika izvršenja događaja kako bi se predvidelo ponašanje sistema, odnosno predvidele aktivnosti i vreme njihovog izvršavanja.



Slika 1. Osnovni koncepti automatskog otkrivanja poslovnih procesa [2]

Prva istraživanja automatskog otkrivanja poslovnih procesa u realnim sistemima su sprovedena u oblasti medicine, praćena sa IKT industrijom, finansijskim sektorom, proizvodnom i obrazovnom industrijom [4]. Mnoge organizacije su prvenstveno koristile automatsko otkrivanje poslovnih procesa za analizu pojedinačnih značajnih procesa unutar organizacija, poput procesa prodaje, nabavke, logistike i proizvodnje. Vremenom je primena poprimila integracijski pristup gde organizacije implementiraju automatsko otkrivanje poslovnih procesa radi kontinuirane analize svih poslovnih procesa [5]. U poslednje dve decenije, pored tehnika automatskog otkrivanja poslovnih procesa, razvijali su se i softverski alati i širila se mogućnost primene automatskog otkrivanja poslovnih procesa u različitim industrijama. Trenutno postoji preko 40 komercijalnih softverskih rešenja koji podržavaju automatsko otkrivanje poslovnih procesa [6], uz dobro poznata rešenja „otvorenog koda” (engl. *open-source*). Primena u industriji se takođe značajno ubrzala u poslednjih 5 godina i očekuje se da će tržište automatskog otkrivanja poslovnih procesa da se duplira na svakih 18 meseci u narednim godinama [6].

1.1. Motivacija, definisanje i opis predmeta istraživanja

Raznoversnost primene automatskog otkrivanja poslovnih procesa i sama činjenica da postoji velik broj različitih alata, tipova primene, algoritama i tehnika, kao i različitih industrija, tipova poslovnih procesa i ciljeva analize dovode do poteškoća prilikom planiranja projekata automatskog otkrivanja poslovnih procesa kod organizacija ili istraživača koji nisu eksperti u primeni. Takođe, s obzirom da je automatsko otkrivanje poslovnih procesa oblast koja se i dalje intenzivno razvija i ne postoji mnogo eksperata u njenoj primeni, doprinosi u pogledu metodologija i uputstava za primenu široke palete dostupnih tehnika su od izuzetnog značaja za praktikante i istraživače iz oblasti. Postoji nekoliko ključnih problema u primeni automatskog otkrivanja poslovnih procesa definisanih od strane Instituta inženjera elektrotehnike i elektronike (engl. *Institute of Electrical and Electronics Engineers – IEEE*) koji će biti razmotreni prilikom definisanja istraživačkog problema.

IEEE radna grupa za automatsko otkrivanje poslovnih procesa je definisala jedanaest ključnih problema odnosno izazova sa kojima se ova oblast susreće prilikom primene u različitim industrijskim sistemima i ka čijem rešavanju naučna zajednica u saradnji sa praktikantima treba da teži [1]. Stoga, doprinos doktorske teze je fokusiran na sledeće probleme iz oblasti [1]:

- Problem 1 (P1) – Čišćenje dnevnika izvršenja događaja;

- Problem 2 (P2) – Benčmarking;
- Problem 3 (P3) – Unapređenje upotrebljivosti za početnike;
- Problem 4 (P4) – Unapređenje razumljivosti za početnike.

Potreba za visoko kvalitetnim podacima je bila očigledna prilikom formalnog definisanja automatskog otkrivanja poslovnih procesa kao oblasti. Jedan od početnih koraka projekata automatskog otkrivanja poslovnih procesa je ekstrakcija, transformacija i čišćenje dnevnika izvršenja događaja. To je ujedno i najzahtevniji korak projekta od kojeg zavise celokupni rezultati analize. Problem **P1 – Čišćenje dnevnika izvršenja događaja**, zahtevao je pre svega definisanje različitih problema kvaliteta koji se mogu javiti u dnevniku izvršenja događaja. Podaci mogu biti ocenjeni u odnosu na njihov stepen zrelosti (engl. *maturity level*), gde stepeni variraju od podataka niskog kvaliteta (npr. podaci ne odgovaraju stvarnosti ili podaci nedostaju) do podataka visokog kvaliteta (npr. podaci su verodostojni i kompletni). Vremenom, kako je automatsko otkrivanje poslovnih procesa sazrevalo kao disciplina i istraživači su zaključili da određeni problemi kvaliteta direktno utiču na celokupan kvalitet analize, javila se potreba za detaljnijom kategorizacijom problema. Bose i dr. [7] su prvi definisali kategorije problema kvaliteta dnevnika izvršenja događaja podelom na nedostajuće podatke (engl. *missing data*), netačne podatke (engl. *incorrect data*), neprecizne podatke (engl. *imprecise data*) i irelevantne podatke (engl. *irrelevant data*). Svaka kategorija problema se može manifestovati kroz različite entitete dnevnika izvršenja događaja, kao što su instanca procesa (engl. *case*), događaj (engl. *event*), naziv aktivnosti (engl. *activity name*), vremenska odrednica aktivnosti (engl. *timestamp*), veza između instance procesa i događaja i resursi procesa. Predstavljena kategorizacije problema kvaliteta dnevnika izvršenja događaja je dalje proširena sa uvođenjem pojma obrazac imperfekcije [8]. Obrasci imperfekcije posmatraju manifestaciju problema kvaliteta podataka kroz različite entitete dnevnika kao obrasce, gde je za svaki obrazac imperfekcije definisana instrukcija za otkrivanje problema i njegovo umanjivanje ili otklanjanje. Ipak, rešenja za otklanjanje problema opisuju kako manualno otkloniti problem bez primene savremenih alata i tehnika za čišćenje dnevnika izvršenja događaja.

Dalja istraživanja su pokušavala da razviju okvire automatizacije procesa detekcije problema kvaliteta podataka [9–13], sa veoma malo pažnje posvećene konkretnim tehnikama čišćenja dnevnika izvršenja događaja koji mogu da umanje ili reše navedene probleme. Jedan sistematski pregled literature tehnika čišćenja dnevnika izvršenja događaja je pokušao da doprinese problemu [14]. Autori su pregledali 70 radova na temu primene tehnika čišćenja dnevnika izvršenja događaja u stvarnim scenarijima i podelili ih u dve kategorije, transformacione tehnike i tehnike detekcije i vizualizacije. Ipak, nisu date konkretne smernice kako izabrati tehniku čišćenja u odnosu na problem kvaliteta dnevnika izvršenja događaja.

Problemi **P2 – Benčmarking**, **P3 - Unapređenje upotrebljivosti za početnike** i **P4 - Unapređenje razumljivosti za početnike** su povezani i mogu se istovremeno diskutovati. Automatsko otkrivanje poslovnih procesa je mlada disciplina i stoga dobar benčmarking i dalje ima prostora za razvoj. Postoji nekoliko desetina tehnika za otkrivanje poslovnih procesa, analizu performansi i proveru usaglašenosti procesa i dnevnika podataka, ali ne postoji konsenzus o kvalitetu i načinu primene ovih tehnika [1]. Problemi unapređenja razumljivosti i upotrebljivosti u oblasti se trenutno rešavaju sa razvojem alata koji će imati intuitivni korisnički interfejs i biti pogodniji za korišćenje. Međutim, trenutno postoji nedovoljno benčmarkinga, ili konsolidacije u pogledu odabira različitih metodologija, alata, algoritama i tehnika za izvođenje različitih koraka projekata automatskog otkrivanja poslovnih procesa, što korisnicima otežava analizu poslovnih procesa i samim tim upotrebljivost automatskog otkrivanja poslovnih procesa. Nastojeći da doprinese rešenju ovog problema, neki autori su predstavili metodologije za upravljanje projektima automatskog otkrivanja poslovnih procesa. Metodologije se mogu podeliti na generičke metodologije i one razvijene isključivo za primenu u zdravstvenoj industriji. Generičke metodologije za primenu AOPP-a u industriji definišu životni ciklus projekata i moguće faze izvršavanje ovih projekata, bez osvrta na aktivnosti nižeg nivoa detaljnosti i odabir algoritama/tehnika koji bi trebali da se primene u svakoj fazi izvršavanje projekta [15–20]. Metodologije razvijene za zdravstvenoj industriju detaljno objašnjavaju kako sprovesti aktivnosti projekta različite granularnosti, kao i koje algoritme i tehnike primeniti na dnevnik izvršenja događaja u različitim situacijama [21–25]. Međutim, ove metodologije i uputstva su primenljivi isključivo na poslovnim procesima i dnevnicima

izvršenja događaja zdravstvene industrije, koja koristi specifične softverske alate i ima specifične ciljeve.

Bitno je napomenuti da su problemi benčmarkinga, unapređenja razumljivosti i upotrebljivosti povezani sa problemom čišćenja dnevnika izvršenja događaja tako što i čišćenje dnevnika izvršenja događaja podrazumeva različite algoritme i tehnike, o čijoj upotrebljivosti i selekciji ne postoji konsolidacija u naučnoj zajednici i među praktikantima automatskog otkrivanja poslovnih procesa.

Na osnovu definisanih problema, predmet istraživanja doktorske disertacije je ispitivanje veze između karakteristika projekata automatskog otkrivanja poslovnih procesa i odabira tipova, algoritama, alata i tehnika automatskog otkrivanja poslovnih procesa, kao i ispitivanje veze između kvaliteta podataka dnevnika izvršenja događaja i odabira tehnika čišćenja dnevnika izvršenja događaja. Iz tog razloga, istraživanje je dvojako, gde se prvi deo istraživanja sastoji iz sistematskog pregleda literature studija slučaja sprovedenih u oblasti automatskog otkrivanja poslovnih procesa i statističkom analizom dobijenih podataka, kako bi se utvrdili ključni faktori primene automatskog otkrivanja poslovnih procesa u različitim industrijama. Drugi deo istraživanja se bavi definisanjem ključnih faktora primene tehnika čišćenja dnevnika izvršenja događaja u zavisnosti od problema kvaliteta podataka, na osnovu ispitivanja praktikanata i istraživača iz oblasti putem upitnika.

1.2. Cilj istraživanja, istraživačka pitanja i hipoteze

Na osnovu prethodno navedenog, cilj istraživanja jeste pružiti sintezu podataka o primeni automatskog otkrivanja poslovnih procesa u različitim industrijskim sistemima, kao i analizu problema kvaliteta podataka i mogućih rešenja iz prakse, kako bi se odredili ključni faktori primene i praktikantima pružila pomoć prilikom donošenja odluke o načinu izvođenja projekata automatskog otkrivanja poslovnih procesa. Istraživanje se dalje može definisati kroz dva istraživačka pitanja (IP):

IP1: Koji faktori projekata analize i automatskog otkrivanja poslovnih procesa (projektne karakteristike) utiču na odabir tipova i algoritama/tehnika automatskog otkrivanja poslovnih procesa?

IP2: Koje tehnike čišćenja treba da se primene na dnevnik izvršenja događaja u zavisnosti od problema kvaliteta podataka?

Na osnovu predstavljenih problema i ciljeva istraživanja, kao i istraživačkih pitanja, definisana su dva konkurentna istraživačka modela. Ključni činioci istraživanja su konstrukti istraživačkih modela i oni su definisani na osnovu izvršenih pregleda literature navedenih u potpoglavlju 1.3.

Prvi istraživački model ima za cilj odgovaranje na IP1 i kroz njega je definisano sledećih 5 hipoteza:

H1: Vrsta industrije ima povezanost sa vrstom poslovnog procesa koji se analizira putem automatskog otkrivanja poslovnih procesa;

H2: Vrsta industrije ima povezanost sa ciljem analize automatskog otkrivanja poslovnih procesa.

H3: Cilj analize automatskog otkrivanja poslovnih procesa ima povezanost sa vrstom poslovnog procesa koji se analizira;

H4: Cilj analize automatskog otkrivanja poslovnih procesa ima povezanost sa tipom automatskog otkrivanja poslovnih procesa;

H5: Cilj analize automatskog otkrivanja poslovnih procesa ima povezanost sa primenjenim algoritmima i tehnikama.

Drugi istraživački model sadrži dimenzije koje se bave IP2 i definiše sledeću hipotezu:

H6: Moguće je odrediti pogodnu tehniku čišćenja dnevnika izvršenja događaja u zavisnosti od problema kvaliteta podataka.

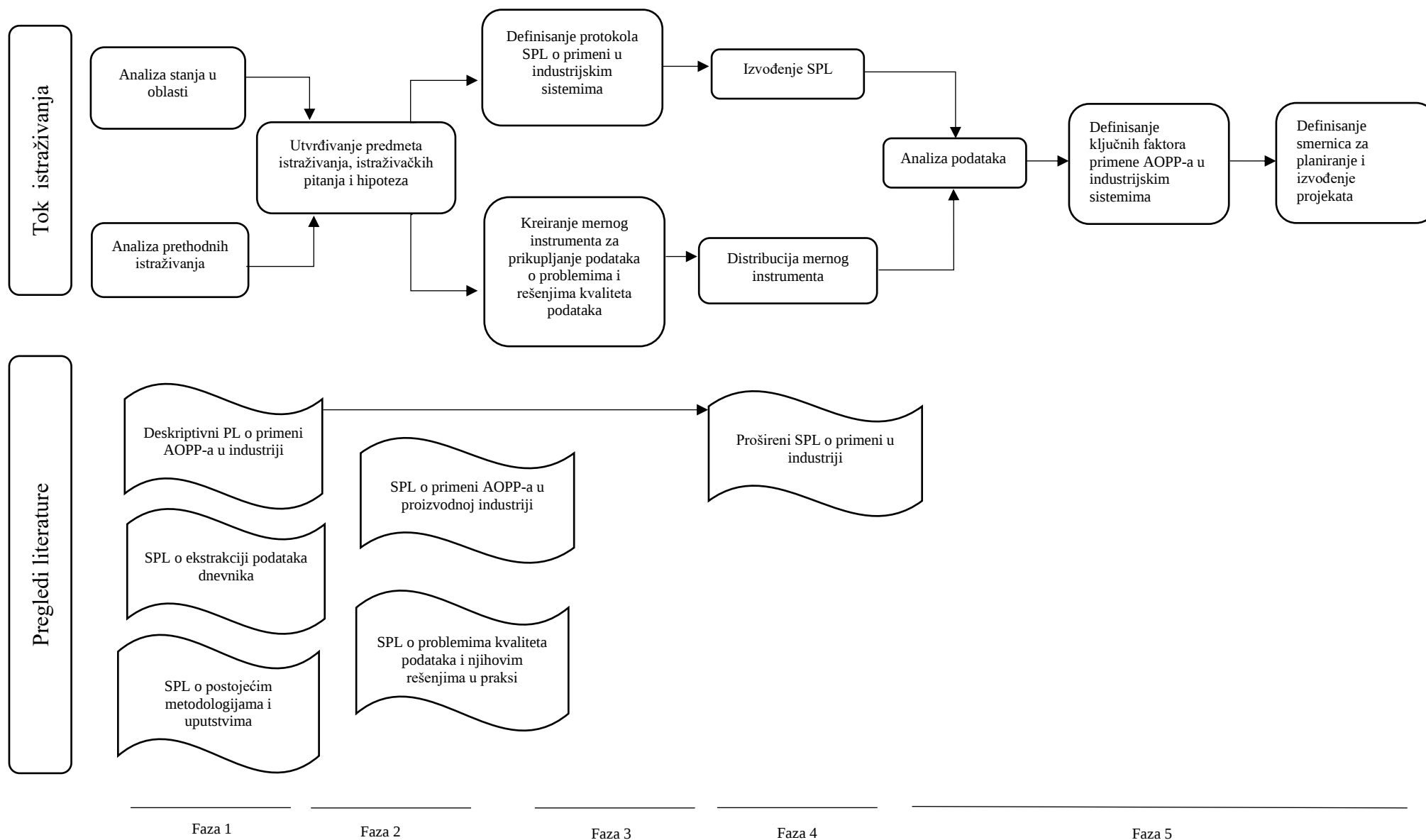
1.3. Tok istraživanja doktorske disertacije

Tok istraživanja predstavljen je na slici 2. Kako bi se utvrdio istraživački problem i istraživačka pitanja, autor je sproveo nekoliko značajnih pregleda literature. Na samom početku, u okviru prve faze istraživanja, izvršen je pregled stanja u oblasti automatskog otkrivanja poslovnih procesa, gde se autor upoznao sa osnovnim konceptima, trendovima i problemima u oblasti. Kao rezultat tog eksplorativnog istraživanja oblasti, objavljen je deskriptivni pregled literature na *Danube Adria Association For Automation & Manufacturing* – DAAAM [4], internacionalnom simpozijumu. Na osnovu inicijalnog istraživanja literature utvrđen je pravac daljeg istraživanja ka problemima upotrebljivosti automatskog otkrivanja poslovnih procesa i problemima kvaliteta podataka, na osnovu čega je bilo moguće izvršiti analizu prethodnih istraživanja na tu temu. Analiza prethodnih istraživanja je uključivala sistematski pregled literature – SPL, načina pribavljanja podataka radi formiranja dnevnika izvršenja događaja, čiji su rezultati takođe objavljeni na internacionalnom simpozijumu iz menadžmenta [26], kao i pregled postojećih metodologija i uputstava za primenu automatskog otkrivanja poslovnih procesa u industriji, kako bi se utvrdila potreba za istraživanjem.

Druga faza istraživanja obuhvata jasno definisanje predmeta istraživanja, istraživačkih pitanja i hipoteza. Postavljanje hipoteza i definisanje istraživačkih modela zahteva jasno utvrđivanje ključnih činilaca istraživanja. Stoga su u ovoj fazi istraživanja sprovedena još 2 sistematska pregleda literature. Prvi SPL ispituje primenu automatskog otkrivanja poslovnih procesa u proizvodnoj industriji, kako bi se na manjem uzorku testirao istraživački model i hipoteze koje se bave IP1. Rezultati istraživanja su objavljeni na *Advances in Production Management Systems* – APMS konferenciji [27], a zatim i u međunarodnom časopisu [28]. Drugi SPL se odnosio na pregled problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja koji se u praksi primenjuju kako bi se odgovarajući problemi rešili. Rezultati pregleda literature su direktno uticali na definisanje mernog instrumenta za IP2 i objavljeni su zajedno sa rezultatima istraživanja u međunarodnom časopisu izuzetnih vrednosti [29].

Treća faza se nadovezuje na rezultate pregledane literature iz prethodne faze i tok istraživanja se dalje deli na dva dela, gde se prvi tok bavi aktivnostima vezanim za IP1, a drugi tok aktivnostima vezanim za IP2. Što se tiče IP1, definišu se procedura sprovođenja sistematskog pregleda literature kojim će se prikupiti sveobuhvatni podaci o ključnim faktorima primene automatskog otkrivanja poslovnih procesa u praksi. Drugi tok treće faze se odnosi na definisanje mernog instrumenta kojim će se ispitati praktikanti automatskog otkrivanja poslovnih procesa o učestalosti pojave problema kvaliteta podataka dnevnika u praksi i mogućim rešenjima.

Četvrta faza obuhvata aktivnosti prikupljanja podataka, kroz sprovođenje sistematskog pregleda literature i distribuciju upitnika. Poslednja, peta faza obuhvata obradu i analizu prikupljenih podataka, pružanje odgovora na istraživačka pitanja, kao i definisanje smernica za primenu automatskog otkrivanja poslovnih procesa u industrijskim sistemima.



Slika 2. Tok istraživanja doktorske disertacije

1.4. Rezultati istraživanja i naučni doprinos disertacije

Doprinos disertacije se pre svega može posmatrati kao teorijski doprinos oblasti automatskog otkrivanja poslovnih procesa, koji obuhvata preglede literature sprovedene u prvoj i drugoj fazi istraživanja, kao i SPL koji je izvršen radi prikupljanja podataka u okviru prvog istraživačkog modela, za potrebe ispitivanja IP1. Spomenuti pregledi literature pružaju uvid u način primene automatskog otkrivanja poslovnih procesa u realnim industrijskim sistemima, kao i učestalost pojave problema kvaliteta podataka unutar dnevnika izvršenja događaja i primene određenih tehnika čišćenja radi umanjivanja ili uklanjanja konkretnog problema kvaliteta.

Pored teorijskog doprinosa, praktični doprinos se može posmatrati kroz rezultate istraživanja sprovedenih u odnosu na dva prethodno definisana istraživačka pitanja.

Prvi istraživački model je težio da utvrdi koji faktori projekata analize i automatskog otkrivanja poslovnih procesa (projektne karakteristike) utiču na odabir tipova i algoritama/tehnika automatskog otkrivanja poslovnih procesa. Na osnovu prethodno izvršenog pregleda literature utvrđena je procedura sistematskog pregleda literature, koji obuhvata 144 primarne studije koje izveštavaju o primeni automatskog otkrivanja poslovnih procesa u realnim industrijskim sistemima. Strategija ekstrakcije podataka je definisana tako da se iz svake primarne studije uključene u pregled literature izvlače podaci o ključnim činiocima ovog istraživačkog modela, a to su: industrija, poslovni proces, cilj analize, tip automatskog otkrivanja poslovnih procesa, kao i specifične primenjene tehnike i softverski alati. Podaci su zatim analizirani kako bi se ispitala postavljene hipoteze, tako što je primenjen Pirsonov Hi-kvadrat test i Fišerov egzaktni test kako bi se ispitala povezanost kategoričkih varijabli, i pritom formirala tabela kontingencije postupkom unakrsne analize, radi sumiranja distribucije povezanosti između posmatranih varijabli. Rezultati pokazuju aktuelno stanje u oblasti primene automatskog otkrivanja poslovnih procesa u pogledu ključnih faktora istraživanja, kao i smernice za odabir tipa i tehnika automatskog otkrivanja poslovnih procesa u odnosu na scenario u okviru kojeg se analiza odvija, određenog vrstom industrije i ciljem analize.

Kako bi se odredilo koje tehnike čišćenja treba da se primene na dnevnik izvršenja događaja u zavisnosti od problema kvaliteta podataka, na osnovu unapred sprovedenih pregleda literature, definisan je upitnik kao merni instrument i distribuiran praktikantima i istraživačima iz oblasti automatskog otkrivanja poslovnih procesa u celom svetu. Upitnik je sadržao stavke koje se tiču učestalosti pojavljivanja određenih problema kvaliteta dnevnika izvršenja događaja i tehnika čišćenja dnevnika izvršenja događaja, kao i njihove posmatrane značajnosti. Takođe, ispitanici su pružali uvid o odabiru tehnike čišćenja dnevnika u zavisnosti od problema kvaliteta dnevnika. Rezultati obrade podataka prikupljenih putem upitnika su pružili značajan doprinos samom stanju u oblasti koje se tiče trendova na tržištu rada i demografskih karakteristika istraživača i praktikanata koji se automatskim otkrivanjem poslovnih procesa.

Takođe, prilikom analize osnovnih demografskih karakteristika ispitanika, izvršen je Hi-kvadrat test i utvrđene su značajne razlike odnosa uloge ispitanika i vremenskog perioda izučavanja/primene automatskog otkrivanja poslovnih procesa, odnosa uloge ispitanika i softverskih alata koje primenjuju za izvođenje automatskog otkrivanja poslovnih procesa, kao i odnosa uloge ispitanika i softverskih alata koje primenjuju za obradu podataka u kontekstu automatskog otkrivanja poslovnih procesa. Zatim su u fokusu analize bili sami problemi kvaliteta podataka i tehnike čišćenja dnevnika izvršenja događaja. Pored osnovne deskriptivne statistike koja pokazuje učestalost pojavljivanja i percipirani značaj ovih varijabli od strane ispitanika, primenjena je i analiza odnosa između značaja i učestalosti (engl. *Importance-Performance Analysis* – IPA).

IPA analiza, na osnovu percepcije ispitanika o učestalosti pojave i značaja posmatranih dimenzija, omogućava uvid u to koji problemi kvaliteta dnevnika podataka i tehnike čišćenja podataka treba da budu u fokusu istraživanja i zahtevaju posebnu pažnju, a koji nisu toliko značajni.

Takođe, primenjena je analiza varijansi (engl. *Analysis of Variance* – ANOVA), koja je pokazala da postoji razlika u značaju koji ispitanici različitih uloga pridaju problemima kvaliteta podataka i

tehnikama čišćenja dnevnika izvršenja događaja, kao i da postoji razlika u pogledu učestalosti sa kojom se ispitanici različitih uloga susreću sa problemima kvaliteta i primenjuju tehnike čišćenja dnevnika izvršenja događaja. Takođe, analiza varijansi je pokazala i da postoji razlika u značaju koji ispitanici različitog nivoa iskustva u oblasti automatskog otkrivanja poslovnih procesa pridaju problemima kvaliteta podataka i tehnikama čišćenja dnevnika izvršenja događaja.

Kako bi se testirala postavljena hipoteza o zavisnosti između problema kvaliteta i tehnika čišćenja dnevnika izvršenja događaja, sproveden je Pirsonov test korelacije, Pirsonov Hi-kvadrat test i unakrsna analiza za definisanje smernica. Rezultati testa korelacije i Hi-kvadrat testa pokazuju da postoji statistički značajna veza između problema kvaliteta i tehnika čišćenja dnevnika izvršenja događaja. Stoga su podaci dobijeni unakrsnom analizom, uz podatke o jačini veze između stavki posmatranih dimenzija dobijenih ispitivanjem korelacije, uzeti u obzir prilikom kreiranja smernica za odabir tehnike čišćenja dnevnika izvršenja događaja.

Smernice za planiranje izvođenja projekata automatskog otkrivanja poslovnih procesa u zavisnosti od vrste industrije i cilja analize automatskog otkrivanja poslovnih procesa, kao i smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu, praktikantima pružaju pomoć prilikom izvršavanja faze planiranja projekata automatskog otkrivanja poslovnih procesa i faze pripreme podataka za analizu putem tehnika automatskog otkrivanja poslovnih procesa [15,16,30].

1.5. Struktura disertacije

Kroz realizaciju faza istraživanja predstavljenih u toku istraživanja doktorske disertacije, oblikovana je struktura disertacije sačinjena od sedam poglavlja koja su objašnjena u nastavku.

Poglavlje 1 – Uvodna razmatranja, prvenstveno opisuje ulogu i značaj automatskog otkrivanja poslovnih procesa, kao oblasti koja omogućava organizacijama da otkriju način izvršavanja poslovnih procesa automatski generišući modele poslovnih procesa na osnovu podataka o izvršavanju aktivnosti koje čine neki poslovni proces. Raznovrsnost primene automatskog otkrivanja poslovnih procesa i sama činjenica da postoji veliki broj dostupnih softverskih alata i algoritama, dovodi do poteškoća prilikom planiranja projekata automatskog otkrivanja poslovnih procesa kod organizacija ili istraživača koji nisu eksperti u oblasti. Istraživanje u okviru ove disertacije motivisano je važnošću otkrivanja ključnih faktora koji utiču na planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa i definisanja odgovarajućih smernica za istraživače i praktikante. U okviru ovog poglavlja predstavljeni su motivacija i predmet istraživanja, cilj istraživanja, istraživačka pitanja i hipoteze, tok istraživanja doktorske disertacije, kao i sažeti rezultati istraživanja.

Poglavlje 2 – Teorijske osnove, struktuirano je tako da uvođenjem neophodnih koncepata i pojmova omogući razumevanje istraživanja sprovedenog u disertaciji. Predstavljeni su osnovni koncepti automatskog otkrivanja poslovnih procesa, koji uključuju definisanje pojmova automatskog otkrivanja poslovnih procesa, dnevnika izvršenja događaja i tipova automatskog otkrivanja poslovnih procesa. Zatim su kroz pregled stanja u oblasti predstavljeni rezultati sistematskih pregleda literature na temu primene automatskog otkrivanja poslovnih procesa u industrijskim sistemima i pregleda problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja koji se primenjuju u praksi. Rezultati prethodnih istraživanja predstavili su postojeće metodologije i uputstva za izvođenje automatskog otkrivanja poslovnih procesa, kao i prethodne sistematske preglede literature u oblasti. Na kraju poglavlja predstavljena su dva istraživačka modela sa istraživačkim hipotezama, nastalim kao rezultat pregleda relevantne literature objašnjene u okviru ovog poglavlja.

Poglavlje 3 – Metodološki aspekti istraživanja objašnjava način na koji su prikupljeni podaci za potrebe istraživanja i podeljeno je na dva dela u skladu sa dva definisana istraživačka modela. Za prvi istraživački model, za koji se podaci prikupljaju putem sistematskog pregleda literature, definisani su protokol i izvođenje sistematskog pregleda literature, zajedno sa predstavljenim bibliografskim podacima primarnih studija koje su uključene u istraživanje. Takođe, navedene su i primenjene metode statističke obrade podataka. Za drugi istraživački model, za koji se podaci prikupljaju putem upitnika,

predstavljen je postupak razvoja mernog instrumenta i distribucije upitnika, kao i detaljan prikaz demografskih karakteristika učesnika u istraživanju i primenjene metode statističke obrade podataka.

Poglavlje 4 – Rezultati istraživanja prikazuje rezultate istraživanja sprovedenog u okviru disertacije, podeljeno na dve celine u skladu sa dva istraživačka modela. Za prvi istraživački model, predstavljeni su deskriptivna statistika varijabli istraživanja, rezultati ispitivanja hipoteza istraživačkog modela 1 i tabela kontingencije odnosa između ključnih faktora istraživanja, na osnovu koje su kasnije definisane smernice za planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa u zavisnosti od vrste industrije i cilja projekta. Za drugi istraživački model predstavljeni su rezultati percipiranog značaja i učestalosti susreta sa problemima kvaliteta podataka, percipiranog značaja i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja, kao i rezultati analize odnosa značaja i učestalosti posmatranih dimenzija istraživanja. Takođe, predstavljeni su i rezultati analize varijansi sprovedene radi ispitivanja razlika između ispitanika u pogledu njihove uloge i iskustva u oblasti automatskog otkrivanja poslovnih procesa. Konačno, predstavljen je i rezultat ispitivanja hipoteze istraživačkog modela 2, na osnovu kojeg su kasnije definisane smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu.

Poglavlje 5 – Diskusija rezultata istraživanja daje osvrt na izvođenje sistematskog pregleda literature i konstrukciju upitnika kao mernog instrumenta i detaljno analizira i objašnjava dobijene rezultate. Ovo poglavlje je strukturirano kroz dva dela u skladu sa diskusijom rezultata istraživačkog modela 1 i istraživačkog modela 2. Za prvi istraživački model diskutuje se o proceduri prikupljanja podataka, deskriptivnoj statistici varijabli istraživanja, rezultatima ispitivanja hipoteza i definišu se smernice za planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa u zavisnosti od vrste industrije i cilja projekta. Za drugi istraživački model diskutuje se o razvoju mernog instrumenta i demografiji ispitanika, analizi značaja i učestalosti susreta sa problemima kvaliteta i tehnikama čišćenja dnevnika izvršenja događaja i rezultatima ispitivanja hipoteze istraživačkog modela 2. Na kraju, predstavljene su definisane smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu.

Poglavlje 6 – Zaključna razmatranja i pravci daljih istraživanja, sumira rezultate ostvarene istraživanjem u disertaciji davanjem odgovora na istraživačka pitanja, navodi teorijske i praktične implikacije rezultata istraživanja, te na osnovu njih objašnjava identifikovane pravce za buduća istraživanja.

Poglavlje 7 – Literatura prikazuje spisak svih naučnih i stručnih radova, istraživanja i drugih relevantnih resursa koji su korišćeni tokom rada na ovoj disertaciji.

Na kraju doktorske disertacije priloženi su bibliografija primarnih studija i upitnik.

2. TEORIJSKE OSNOVE

Ovo poglavlje predstavlja teorijske podloge sprovedenog istraživanja doktorske disertacije. Sačinjeno je iz više potpoglavlja, gde se prvenstveno definišu osnovni koncepti automatskog otkrivanja poslovnih procesa, čije je poznavanje neophodno radi razumevanja metodologije i rezultata istraživanja. Zatim sledi pregled stanja u oblasti, koji prikazuje rezultate sistematskih pregleda literature sprovedenih radi utvrđivanja ključnih faktora istraživanja. Radi utvrđivanja potrebe za istraživanjem, predstavljeni su rezultati prethodnih istraživanja koji uključuju postojeće metodologije i uputstva za automatsko otkrivanje poslovnih procesa, kao i prethodne preglede literature iz oblasti. Zatim su predstavljena dva istraživačka modela sa hipotezama, definisana na osnovu teorijskih osnova.

2.1. Osnovni koncepti automatskog otkrivanja poslovnih procesa

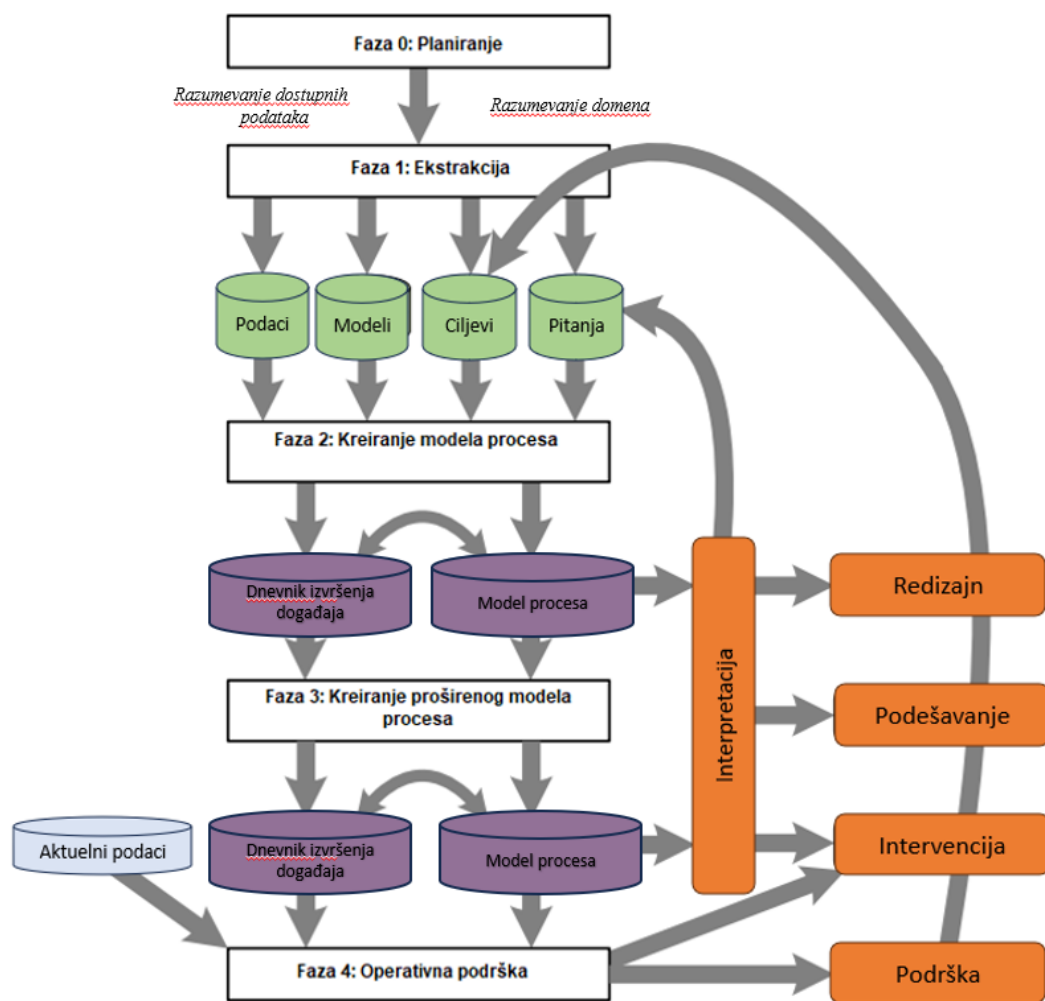
U ovom potpoglavlju je pre svega objašnjen sam pojam automatskog otkrivanja poslovnih procesa, uz prikaz životnog ciklusa projekta automatskog otkrivanja poslovnih procesa. Zatim je pružen uvid u definiciju i strukturu dnevnika izvršenja događaja. Na kraju, jasno su definisani svi tipovi automatskog otkrivanja poslovnih procesa uz navedene najčešće primenjivane algoritme i tehnike.

2.1.1. Definicija automatskog otkrivanja poslovnih procesa

IEEE je osnovao radnu grupu za podršku oblasti automatskog otkrivanja poslovnih procesa (engl. *The IEEE Task Force on Process Mining*) koja je 2011. definisala oblast, osnovne koncepte i probleme kroz manifest [1]. Prema njima, definicija automatskog otkrivanja poslovnih procesa je sledeća [1]:

„Ideja automatskog otkrivanja poslovnih procesa je otkrivanje, praćenje i unapređenje postojećih poslovnih procesa, tako što se iz savremenih informacionih sistema prikuplja znanje o izvršenju poslovnih procesa. Automatsko otkrivanje poslovnih procesa podrazumeva ekstrakciju modela procesa iz dnevnika izvršenja događaja, proveravanje usaglašenosti poslovnog procesa odnosno posmatranje devijacija kroz poređenje modela procesa i dnevnika izvršenja događaja, zatim analizu društvenih mreža, automatsko definisanje simulacionih modela, proširenje modela procesa, popravljavanje modela procesa, predviđanje ponašanja sistema i donošenje preporuka na osnovu istorijskog ponašanja sistema“.

Slika 3 prikazuje životni ciklus jednog projekta automatskog otkrivanja poslovnih procesa, sa konkretnim aktivnostima i specifičnim konceptima [1]. Svaki projekat bi trebao da započne sa fazom planiranja (Faza 0). Nakon iniciranja projekta, faza 1 podrazumeva izvlačenja podataka, definisanje modela procesa na osnovu ugrađenog znanja vlasnika procesa, definisanje istraživačkih pitanja i ključnih indikatora performansi (engl. *Key Performance Indicators* – KPI). Navedene aktivnosti zahtevaju poznavanje dostupnih podataka koji opisuju izvršavanje procesa i razumevanje domena istraživanja, odnosno koja su istraživačka pitanja od značaja za analizu konkretnog procesa. Faza 2 sadrži najbitnije aktivnosti projekta, a to su automatsko definisanje toka izvršavanja procesa na osnovu formiranog dnevnika izvršenja događaja. Otkriveni poslovni proces već može da pruži odgovore na neka od postavljenih istraživačkih pitanja i da pokrene aktivnosti redizajna ili podešavanja aktivnosti. Takođe, moguće je filtrirati dnevnik izvršenja događaja kako bi se prilagodio potrebama istraživanja i ponovo izvršiti aktivnosti iz faze 2. Kada je otkrivanje modela procesa izvedeno u željenom kvalitetu, moguće je proširiti dnevnik izvršenja događaja sa podacima potrebnim za analizu iz drugih perspektiva (analiza performansi, provera usaglašenosti i analiza društvenih mreža). U fazi 3 je stoga moguće proširiti model procesa sa podacima o vremenskim odrednicama i resursima koji su utrošeni prilikom izvođenja aktivnosti. Modeli konstruisani u ovoj fazi mogu pružiti ulazne informacije za operativnu podršku u fazi 4.



Slika 3. Životni ciklus projekta automatskog otkrivanja poslovnih procesa [1]

2.1.2. Dnevnik izvršenja događaja

IEEE radna grupa za podršku oblasti automatskog otkrivanja poslovnih procesa definiše dnevnik izvršenja događaja (engl. *event log*) na sledeći način [1]:

„Sve tehnike automatskog otkrivanja poslovnih procesa pretpostavljaju da je moguće sekvencijalno beležiti podatke o događajima tako da svaki događaj odgovara aktivnosti, odnosno dobro definisanom koraku procesa i svaki događaj pripada određenom slučaju procesa, odnosno instanci izvršenja procesa. Dnevnik izvršenja događaja dalje sadrži podatke o vremenskim odrednicama izvršenja aktivnosti, resursima utrošenim prilikom izvršenja aktivnosti i drugim elementima podataka zabeleženim uz događaje (npr. veličina porudžbine)“.

Tabela 1 predstavlja izvod iz jednog primera dnevnika izvršenja događaja [31] koji se odnosi na proizvodni proces i sadrži minimum informacija neophodnih za sprovođenje svih tipova automatskog otkrivanja poslovnih procesa.

Tabela 1. Primer dnevnika izvršenja događaja [31]

ID Slučaja	Naziv aktivnosti	Vremenska odrednica	Resurs
Slučaj 1	Glodanje	29-01-23 23:24	Mašina 1

Slučaj 1	Lasersko obeležavanje	30-01-23 05:44	Mašina 2
Slučaj 1	Brušenje	30-01-23 06:59	Mašina 3
Slučaj 1	Pakovanje	30-01-23 07:21	Zaposleni 1
Slučaj 2	Glodanje	31-01-23 13:20	Mašina 1
Slučaj 2	Lasersko obeležavanje	01-02-23 08:18	Mašina 2

Kako je dnevnik izvršenja događaja sačinjen od skupa instanci procesa, jedinstveni identifikator (ID) slučaja je neophodan kako bi se upravljalo individualnim instancama procesa i kako bi se pojedinačni događaji povezali sa slučajem procesa unutar kojeg su izvršeni. Svaki slučaj se sastoji od sekvence događaja izvršenih u okviru jedne pojave procesa, gde su događaji određeni kao aktivnosti sa nazivom. Vremenska odrednica je atribut koji opisuje kada je događaj izvršen i omogućava definisanje redosleda izvršenja događaja. Dnevnik izvršenja događaja može da sadrži razne dodatne podatke koji ga proširuju i omogućavaju detaljniju analizu, poput podataka o resursima procesa koji omogućavaju analizu društvenih mreža (engl. *social network analysis*).

Zbog kompleksnih veza između događaja, slučaja pojave procesa i njihovih atributa, razvijen je *eXtensible Event Stream* (XES) standard, kao jezik za prenos, čuvanje i razmenu podataka dnevnika izvršenja događaja, odobren od strane IEEE asocijacije za standarde. XES koristi W3C XML (*World Wide Web Consortium W3C; Extensible Markup Stream XML*) shemu definicije podataka, čime garantuje interoperabilnost među sistemima.

Kao što je spomenuto ranije, autori Bose i dr. [7] i Suriadi i dr. [8] su definisali četiri široke kategorije problema kvaliteta podataka u oblasti automatskog otkrivanja poslovnih procesa (nedostajući podaci, netačni podaci, neprecizni podaci i irelevantni podaci), uz definisane obrasce imperfekcije koji opisuju specifične manifestacije problema kroz entitete dnevnika izvršenja događaja. Ova tema će se detaljno obraditi u potpoglavlju 2.2.3.1. Sinteza problema kvaliteta podataka dnevnika izvršenja događaja.

2.1.3. Tipovi i perspektive automatskog otkrivanja poslovnih procesa

Tipovi i perspektive automatskog otkrivanja poslovnih procesa se mogu podeliti na one koji analiziraju istorijske podatke o izvršavanju procesa i pružaju uvid u to kako se proces odvijao (engl. *backward-looking*) i one koji analiziraju iste te podatke kako bi predvideli ponašanje procesa u budućnosti (engl. *forward-looking*). U prvu grupu tipova spadaju otkrivanje poslovnih procesa, unapređenje poslovnog procesa i provera usaglašenosti procesa, kao i perspektiva analize društvenih mreža (engl. *social network analysis*). U drugu grupu spadaju prediktivne tehnike automatskog otkrivanja poslovnih procesa, koje se često kombinuju sa drugim oblastima nauke o podacima poput mašinskog učenja.

Otkrivanje poslovnog procesa ima za cilj da otkrije model poslovnog procesa na osnovu podataka dnevnika izvršenja događaja i predstavi ga kroz neku od prihvaćenih notacija, poput Petrijeve mreže, BPMN jezika, grafa direktnog praćenja (engl. *Directly-Follows graph* – DFG) i drugih. Postoje četiri kriterijuma koji određuju nivo kvaliteta modela procesa u odnosu na to u kojoj meri otkriveni model tačno predstavlja podatke iz dnevnika izvršenja događaja. Prvi kriterijum je sposobnost modela procesa da opiše ponašanje sistema zabeleženo u dnevniku izvršenja događaja (engl. *replay fitness*). Drugi kriterijum je preciznost modela (engl. *overfitting*), koji teži ka tome da model ne predstavlja ponašanje koje nije zabeleženo u dnevniku izvršenja događaja. Suprotno, treći kriterijum je generalizacija, koji teži ka tome da model generalizuje ponašanje sistema u smislu da ne omogućava samo ponašanje zabeleženo u dnevniku izvršenja događaja. Poslednji kriterijum je jednostavnost otkrivenog modela procesa, gde model mora biti dovoljno razumljiv i čitljiv.

Kako bi se ispunili uslovi koje nalažu kriterijumi kvaliteta modela procesa, razvijeni su različiti algoritmi otkrivanja procesa.

Kao prvi algoritam za otkrivanje poslovnih procesa, alfa (α) algoritam je nastao na Tehničkom Univerzitetu u Ajndhovenu kao jednostavan deterministički algoritam koji je imao mogućnost otkrivanja konkurentnih modela procesa, uključujući koncepte kao što su paralelne i rekurzivne aktivnosti. Alfa

algoritam je posmatrao samo redosled aktivnosti zabeleženih u dnevniku izvršenja događaja i nije uzimao u obzir vremensku odrednicu i ID slučaja. Na osnovu redosleda aktivnosti, kreirao je skup putanja izvršavanja procesa (engl. *trace*) koji sadrži samo nazive poređanih aktivnosti. Cilj α -algoritma je da kreira model predstavljen kroz Petrijevu mrežu koji će opisati sva moguća ponašanja unutar procesa.

Osnovna ideja α -algoritma će biti objašnjena na primeru [3]. Posmatramo dnevnik izvršenja događaja D1, koji sadrži skup putanja izvršavanja procesa sa aktivnostima imenovanim malim latiničnim slovima.

$$D1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle]$$

Algoritam prvo analizira odnose sledljivosti između svake dve pojedinačne aktivnosti unutar dnevnika izvršenja događaja. Mogući odnosi su sledeći:

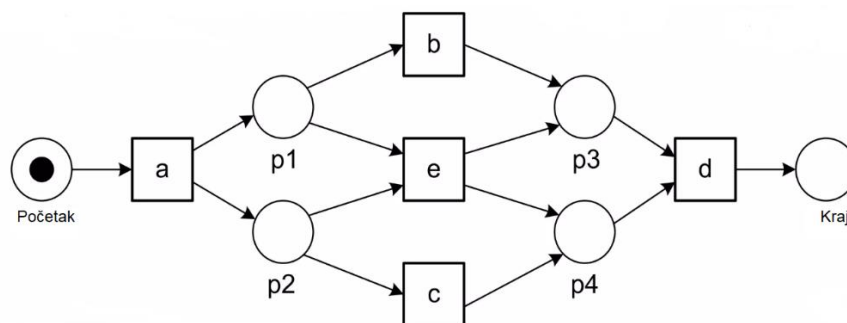
- Direktno nasleđivanje (engl. *direct succession*), se javlja kada jedna aktivnost u bar jednom slučaju direktno sledi drugu aktivnost. Obeležava se sa simbolom „ $>$ “. Na primer, u dnevniku D1 aktivnost a je direktno praćena sa aktivnošću b , stoga $a > b$.
- Kauzalnost (engl. *causality*), se javlja kada jedna aktivnost u bar jednom slučaju direktno sledi drugu aktivnost, ali ne i obrnuto. Obeležava se sa simbolom „ $\rightarrow$ “. Na primer, u dnevniku D1 aktivnost a je direktno praćena sa aktivnošću b , ali aktivnost b se nikada ne odvija nakon aktivnosti a , stoga $a \rightarrow b$.
- Paralelno izvršavanje (engl. *parallel*), se javlja kada jedna aktivnost u bar jednom slučaju direktno sledi drugu aktivnost i obrnuto. Obeležava se sa simbolom „ $\parallel$ “. Na primer, u dnevniku D1 aktivnost b je direktno praćena sa aktivnošću c , i aktivnost c je direktno praćena sa aktivnošću b , stoga $b \parallel c$ i $c \parallel b$.
- Izbor (engl. *choice*), se javlja kada jedna aktivnost ni u jednom slučaju direktno ne sledi drugu aktivnost i obrnuto. Obeležava se sa simbolom „ $\#$ “. Na primer, u dnevniku D1 aktivnost a nije direktno praćena sa aktivnošću d i aktivnost d nije direktno praćena sa aktivnošću a , stoga sledi $a \# d$ i $d \# a$.

Na osnovu otkrivenih odnosa između aktivnost, α -algoritam dalje pravi otisak (engl. *footprint*) dnevnika izvršenja događaja, koji se može predstaviti kao matrica u Tabeli 2.

Tabela 2. Matrica sledljivosti aktivnosti dnevnika D1 – „Otisak“

	a	b	c	d	e
a	#	$\rightarrow$	$\rightarrow$	#	$\rightarrow$
b	$\leftarrow$	#	$\parallel$	$\rightarrow$	#
c	$\leftarrow$	$\parallel$	#	$\rightarrow$	#
d	#	$\leftarrow$	$\leftarrow$	#	$\leftarrow$
e	$\leftarrow$	#	#	$\rightarrow$	#

Podaci iz matrice sledljivosti se koriste kako bi se otkrili obrasci potrebni za modelovanje modela procesa, poput sekvencijanog toka, ekskluzivne kontrole toka (engl. *XOR-split pattern* i *XOR-join pattern*) i paralelne kontrole toka (engl. *AND-split pattern* i *AND-join pattern*). Na osnovu primene α -algoritma na dnevnik D1, automatski je otkriven model procesa predstavljen kao Petrijeva mreža na slici 4.



Slika 4. Petrijeva mreža modela procesa dobijena na osnovu primene α -algoritma na dnevnik D1 [3]

Alfa algoritam je imao dosta nedostataka prilikom primene na kompleksnije dnevnike izvršenja događaja koji su predstavljali procese izvršene u realnim industrijskim sistemima [3,5]. S obzirom na fokusiranost algoritma isključivo na otkrivanje perspektive kontrole toka procesa, nije obuhvatao složenije aspekte procesa gde se određene aktivnosti mogu odvijati različitim redosledom i gde postoji velika fleksibilnost ponašanja sistema. Prilikom obrade velike količine podataka, α -algoritam je sklon nepreciznosti i preteranoj generalizaciji prilikom detektovanja paralelnih aktivnosti, te je moguće da kao rezultat pruži netačne modele procesa. Takođe, iako α -algoritam može otkriti petlje (engl. *loop*), ne može tačno da izmodeluje složenije strukture, poput ugnježđenih petlji.

Iz tih razloga, istraživači su nastavili da razvijaju algoritme otkrivanja modela procesa koje će pružati modele višeg kvaliteta. Heuristički algoritmi traže optimalan model poslovnog procesa prema principu „pokušaja i greške” i koriste se za otkrivanje kompleksnih modela procesa. U njih spadaju dva najčešće primenjivana algoritma, Induktivni (engl. *Inductive miner*) i heuristički (engl. *Heuristic miner*). Sa genetskim algoritmima, rešenje problema se pronalazi započinjanjem sa proizvoljnom početnom tačkom, a zatim traženjem boljeg rešenja uz zanemarivanje inferiornog rešenja, gde se pretraživanje vrši izvođenjem slučajnih varijacija [32]. Uz ovakav pristup, genetski algoritam iterativno unapređuje model poslovnog procesa, dok ne dođe do prihvatljivog rešenja. Napredne tehnike otkrivanja poslovnih procesa imaju mogućnost da u većoj meri izmodeluju kompleksne modele procesa visokog kvaliteta, poput algoritama zasnovanih na teoriji regija (engl. *theory of regions*), algoritma podele (engl. *Split Miner*) i algoritma skeleta dnevnika (engl. *Log Skeleton*) [33].

Neki algoritmi otkrivanja poslovnih procesa imaju mogućnost analize proširenog dnevnika izvršenja događaja, gde radi analize performansi uključuju i podatke o vremenskim odrednicama događaja i utrošenim resursima. U tim slučajevima, radi se o pristupima unapređenja poslovnog procesa. Ovaj tip automatskog otkrivanja poslovnih procesa obuhvata perspektivu analize performansi i ima kao krajnji cilj izmenu ili proširenje inicijalnog modela procesa. Jedna vrsta unapređenja jeste direktna izmena modela procesa kako bi mu se povećao kvalitet. Druga vrsta se fokusira na proširenje modela sa dodatnim podacima o izvršenju procesa i omogućava određivanje vreme trajanja procesa, frekvenciju izvršenja aktivnosti, frekvenciju veza između aktivnosti, određivanje vremena čekanja između aktivnosti i utvrđivanje „uskih grla” (engl. *bottlenecks*) u procesu. Ukoliko dnevnik izvršenja događaja sadrži podatke o resursima, moguće je izvršiti analizu društvenih mreža kako bi se stekao uvid u saradnju između učesnika u procesu i raspodelu posla među njima.

Treći tip automatskog otkrivanja poslovnih procesa je provera usaglašenosti između otkrivenog modela procesa i podataka unutar dnevnika izvršenja događaja. Provera usaglašenosti se zasniva na kvantifikaciji kriterijuma kvaliteta modela procesa koji se tiču mogućnosti modela da se nad njim odigra proces onako kako je zabeležen u dnevniku, kao i merenje preciznosti i generalizacije procesa. Pored određivanja kvaliteta modela procesa, provera usaglašenosti može da se koristi i za pronalaženje i kvantifikaciju devijacija unutar procesa. Neki softverski alati imaju ugrađene tehnike provere usaglašenosti, koje se zasnivaju na tri pristupa povezivanja modela procesa sa posmatranim ponašanjem sistema: provera pravila (engl. *rule-checking*), ponovno izvođenje tokena (engl. *token replay*) i najčešće primenjivan pristup, poravnavanje (engl. *alignment*) [34].

Tri predstavljena tipa automatskog otkrivanja poslovnih procesa predstavljaju standardne tipove utvrđene manifestom, koji posmatraju istorijske podatke o izvršavanju poslovnog procesa [1]. Međutim, integracija tehnika predikcije sa automatskim otkrivanjem poslovnih procesa se u poslednjih nekoliko godina razvila u dovoljnoj meri da se u novijoj literaturi posmatra kao poseban tip automatskog otkrivanja poslovnih procesa [5]. Standardne tehnike mašinskog učenja poput regresije i stabla odlučivanja se mogu primeniti na prilagođene podatke iz proširenog dnevnika izvršenja događaja kako bi se predvidelo ponašanje sistema, npr. pojava uskih grla, predikcija vremena čekanja ili predikcija pojave devijacija.

2.2. Pregled stanja u oblasti

Pregled stanja u oblasti prikazuje rezultate sistematskih pregleda literature sprovedenih radi utvrđivanja ključnih faktora istraživanja, a koji se tiču primene automatskog otkrivanja poslovnih procesa u industrijskim sistemima, primene automatskog otkrivanja poslovnih procesa u proizvodnim sistemima i problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja.

2.2.1. Pregled primene automatskog otkrivanja poslovnih procesa u industrijskim sistemima

Prvi pregled literature koji je sproveden za potrebe ovog istraživanja je imao za cilj da prikupi značajne informacije o primeni automatskog otkrivanja poslovnih procesa u industriji, sa naglaskom na tipove, perspektive, grane industrije i vrste poslovnih procesa koji su analizirani. Takođe, posmatrane su i primenjene tehnike automatskog otkrivanja poslovnih procesa, kao i softverski alati. Detaljna procedura i rezultati analize su objavljeni kao rad na DAAAM konferenciji [4], a u daljem tekstu biće predstavljen sažetak.

Pretraga radova se odvijala unutar *Scopus* indeksne baze. Kriterijum pretrage je definisan na osnovu ključnih termina pretrage:

„*process mining*” AND (*application* OR *industry* OR *case study*) AND („*business process*“ OR „*information systems*“ OR „*business process management*”),

koji se mogu pojaviti u naslovu rada, abstraktu ili ključnim rečima.

Primenjeno je ograničenje da radovi moraju da budu članci iz časopisa ili radovi sa konferencija, objavljeni između 2009. i 2018. godine, napisani na engleskom jeziku. Radovi objavljeni između 2009. i 2013. godine su morali imati minimum 10 citata, kako bi se isključila nerelevantna istraživanja.

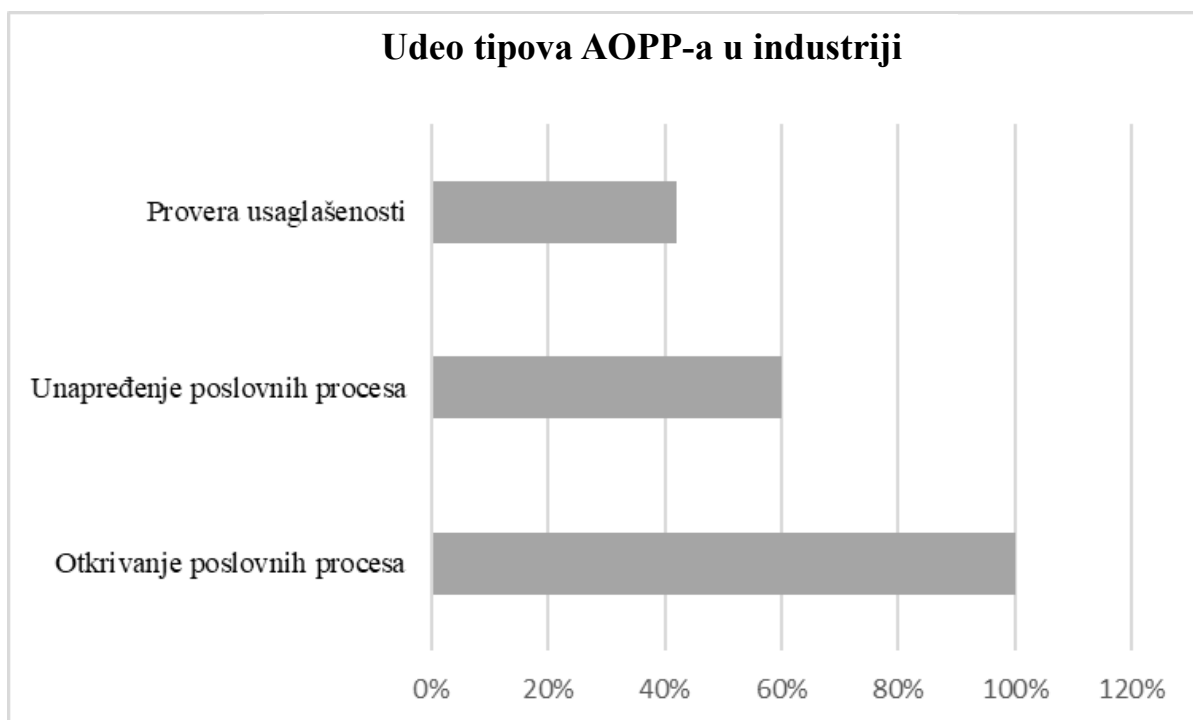
Identifikacija relevantnih radova na osnovu kriterijuma istraživanja rezultirala je sa 187 rezultata pretrage. Relevantni radovi su uvezeni u *Mendeley* softverski alat za upravljanje referencama, gde je obavljen pregled naslova i abstrakata, što je rezultiralo isključenjem 112 radova. U ovoj fazi, radovi su uglavnom bili isključeni jer su se fokusirali samo na temu istraživanja i eksploatacije podataka, a ne automatskog otkrivanja poslovnih procesa, nisu predstavljali praktične primene tehnika automatskog otkrivanja poslovnih procesa u industriji ili su se bavili samo temom problema i rešenja u oblasti automatskog otkrivanja poslovnih procesa. U narednoj fazi pregleda, zbog ograničenog pristupa radovima, iz konačne selekcije je izostavljen određen broj radova, što je rezultiralo 50 dobijenih potencijalnih primarnih studija, pogodnih za sledeću fazu – kritička ocena studija. Nakon kritičke ocene studija, izdvojena su 34 rada koji su prihvaćeni kao relevantni, verodostojni, pronicljivi i dovoljno rigorozni da se uključe u ekstrakciju podataka.

Ekstrakcija podataka je bila fokusirana na primenjene perspektive automatskog otkrivanja poslovnih procesa, tipove automatskog otkrivanja poslovnih procesa, industrije u okviru kojih su analizirani poslovni procesi, tipove poslovnih procesa, kao i korišćene softverske alate i algoritme/tehnike za automatsko otkrivanje poslovnih procesa. Nakon uvoza podataka u *Microsoft Excel*, sprovedena je dalja analiza, fokusirajući se na učestalost pojavljivanja posmatranih koncepata, kao i na otkrivanje značajnih odnosa među njima.

Perspektiva kontrole toka se primenjuje u svim primarnim studijama, što nije neočekivano, jer predstavlja osnovu za dalju analizu procesa [19,24,35–66]. Perspektive slučaja i vremena u većini radova su korišćene zajedno, pa su predstavljene kao jedna perspektiva koja se primenjuje u 63% posmatranih primarnih studija, uglavnom za: analizu vremena protoka [19,20,35,48,50,51,63,65,66], otkrivanje obrazaca i učestalosti slučajeva [20,21,24,37,40,41,47,48,56,57,60,61,63,66], utvrđivanje uskih grla i retkog ponašanja [19,20,24,35,41,57,65], analizu atributa i resursa [35,37,40,44,46,48,61], otkrivanje povratnih petlji [50,56,57] i analiza vremenskih trendova i vremena čekanja [24,40,41,50,57,60,61,66].

Organizaciona perspektiva je korišćena samo u 22% primarnih studija, iako je prema percepciji autora bila jednako korisna kao i druge perspektive. U [52], istraživanja su koristila tehnike otkrivanja društvenih mreža kako bi otkrili strukturu posade glavne kontrolne sobe u nuklearnom postrojenju, na osnovu njihovih komunikacionih sadržaja koji se izgovaraju u toku snalaženja u scenariju obuke. Štaviše, u [55] istraživači su otkrili modele interakcije uloga u hitnoj pomoći, kroz upotrebu kliničkih podataka iz stvarnog života, pružajući relevantne uvide o tome kako zdravstveni radnici sarađuju i otkrivajući mogućnosti za poboljšanje procesa. U [21] fokus je bio na analizi podataka o ponašanju učenika kroz primenu tehnika otkrivanja društvenih mreža posebno usmerenih na složen proces rešavanja problema, sa nalazima koji pokazuju da određeni obrasci ponašanja zaista mogu biti povezani sa boljim/gorim ishodima u smislu postizanja zadovoljavajućeg kvaliteta modela. U [19,63] izvršena je analiza uloga, kako bi se proverila efikasnost podele dužnosti. U [65], istraživači su generisali model procesa timskog toka, gde tok predstavlja niz timova koji su radili na istom dokumentu. U [39] tehnike otkrivanja društvenih mreža su korišćene za mapiranje interakcija koje se dešavaju u procesu i definisanje oblasti sa najvećim brojem interakcija.

Slika 5 prikazuje raspodelu tipova automatskog otkrivanja poslovnih procesa prema učestalosti primene u analiziranim primarnim studijama.

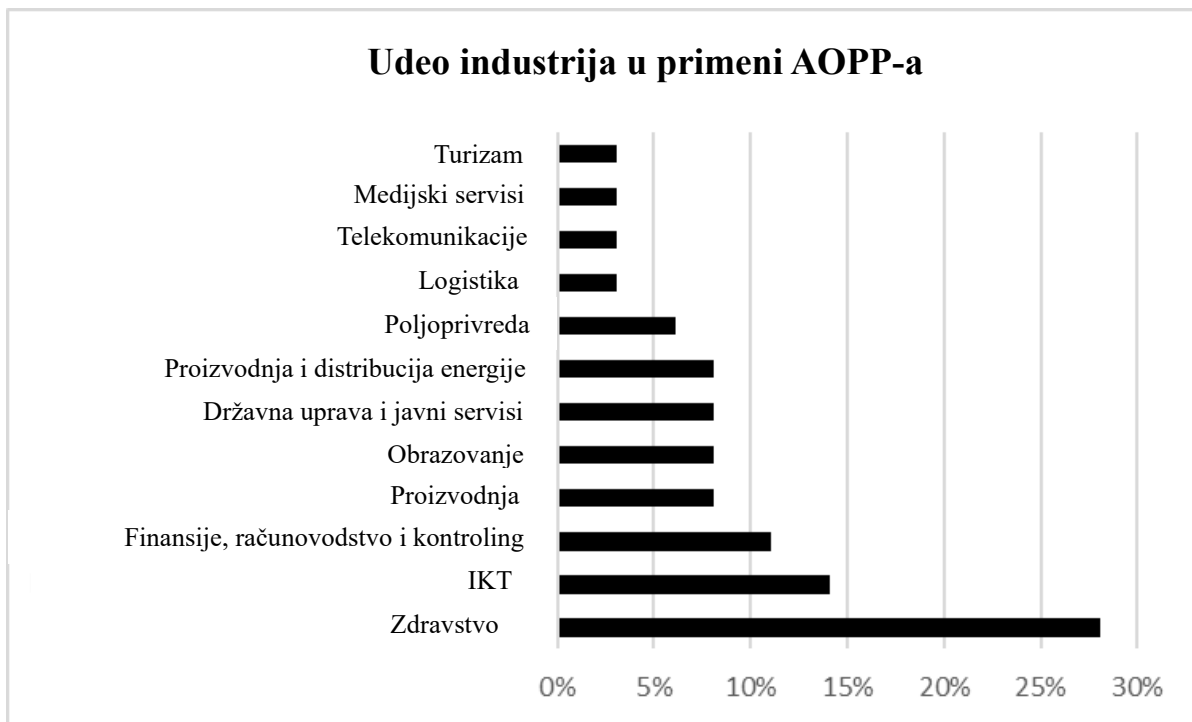


Slika 5. Udeo primene tipova automatskog otkrivanja poslovnih procesa u industriji

Tip otkrivanja poslovnih procesa se u potpunosti poklapa sa perspektivom otkrivanja kontrole toka, jer oba prilaza imaju za cilj definisanje modela procesa. Unapređenje poslovnih procesa je drugi najčešće primenjivani tip, sa udelom od 60%. Potrebno je naglasiti da se perspektiva analize slučaja/vremena sprovodi u svim primarnim studijama gde se izvodi unapređenje poslovnih procesa. Provera usaglašenosti, kao najmanje primenjena vrsta automatskog otkrivanja poslovnih procesa, primenjena je u 42% primarnih studija, što implicira da se istraživači uglavnom oslanjaju na sva tri tipa automatskog otkrivanja poslovnih procesa. Motivacija za primenu provere usaglašenosti, kako se navodi

u radovima, bila je poređenje otkrivenog modela sa postojećim, referentnim modelom, zatim provera da li je realnost usklađena sa modelom i obrnuto, kako bi se otkrila odstupanja i dale sugestije za unapređenje modela procesa. Generisani i postojeći referentni modeli procesa (u nekim slučajevima referentni model procesa su razvili istraživači intervjujući menadžere i relevantno osoblje kompanije) morali su da se kvantifikuju radi direktnog poređenja, stoga su istraživači koristili metriku podobnosti i prikladnosti, uglavnom u softverskom alatu *ProM* (Samo su u [44] istraživači su izabrali softverski alat *Disco*).

Slika 6 prikazuje industrije u posmatranim primarnim studijama u kojima je primenjeno automatsko otkrivanje poslovnih procesa, kao i njihov udeo.



Slika 6. Industrije u posmatranim primarnim studijama i njihov udeo

Zdravstvena industrija u velikoj meri primenjuje automatsko otkrivanje poslovnih procesa, sa udelom od 28%. S obzirom na tipove automatsko otkrivanje poslovnih procesa primenjivane u ovoj industriji, mera u kojoj je primenjeno otkrivanje modela procesa i unapređenje odgovara proseku. Međutim, provera usaglašenosti je obavljena u 20% primarnih studija zdravstvene industrije, naspram 42% primarnih studija iz svih industrija. Za analizu performansi procesa istraživači su koristili *ProM* u 22%, *Disco* u 66% i oba u 11% primarnih studija. Procesi koji su bili tema istraživanja u zdravstvu su: tok kretanja pacijenata (pojavljuje se u 70% primarnih studija), pri čemu se deo primarnih studija [51,55,66,67] fokusira na put pacijenata u hitnoj pomoći, deo [37,39,45] na tok pacijenta uopšte, a tu su i analize ambulantnog procesa [60], korišćenja medicinske mobilne aplikacije [54] i generisanje personalizovanog procesa zdravstvene nege [44].

IKT industrija se detektuje u 14% primarnih studija. Analiza radova sa fokusom na tipove automatskog otkrivanja poslovnih procesa, perspektive i softverske alate pokazala je zanimljive rezultate. Konkretno, provera usaglašenosti je izvršena u 60% primarnih studija, ali što je još iznenađujuće, predlozi za poboljšanje nisu postignuti ni u jednom radu, jer nisu primenjene perspektive slučaja, vremena i organizacione analize. Štaviše, za otkrivanje procesa i perspektive kontrole toka, korišćen je samo softverski alat *ProM*. Automatsko otkrivanje poslovnih procesa se koristi u IKT industriji za: procenu softverskih procesa [53,59], otkrivanje poslovnog procesa ugrađenog u nasleđeni sistem [22], otkrivanje abnormalne aktivnosti korisnika na veb sajtovima društvenih mreža [62] i otkrivanje incidenata i rukovanje zahtevima u upravljanju IT uslugama [38].

Industrija finansija, računovodstva i kontrolinga je identifikovana u 11% primarnih studija, uz pojačano korišćenje organizacione perspektive (50%, u poređenju sa prosekom od 22% za sve primarne

studije), perspektive slučaja/vremena (75%, u poređenju sa prosekom od 63% za sve radove), a samim tim i unapređenje, koje se pojavljuje u 75% primarnih studija u finansijskoj industriji. Procesno rudarenje se koristi u finansijama za: otkrivanje ublažavanja internih transakcionih prevara u procesu nabavke [63], analizu sistema upravljanja dokumentima unutar organizacije koja pruža finansijske usluge [65] i analizu procesa odobravanja kredita u banci [48].

Ostale industrije predstavljene na slici 6 su primenjene u manje od 10% primarnih studija.

Naredni činilac istraživanja koji je posmatran jesu upotrebljeni softverski alati oblasti automatskog otkrivanja poslovnih procesa, zajedno sa konkretnim algoritmima i tehnikama.

Nakon pregleda radova i izdvajanja podataka o softverskim alatima, algoritmima i tehnikama korišćenim za automatsko otkrivanje poslovnih procesa, najvažniji podaci su prikazani u sledećim tabelama. Alati koji se koriste u svakoj primarnoj studiji su *ProM* i *Disco*, pri čemu je *ProM* dominantan alat koji se koristi u 75% primarnih studija, a *Disco* u 44% primarnih studija. U nekoliko radova korišćeni su i *ProM* i *Disco*. *ProM* je trenutno najistaknutiji softverski alat za automatsko otkrivanje poslovnih procesa, koji nudi platformu otvorenog koda sa brojnim dostupnim dodacima, omogućavajući širok obim analize. S druge strane, *Disco* je jednostavan, brz i najefikasniji okvir za upravljanje podacima dnevnika i filtriranje, sa intuitivnim korisničkim interfejsom i vizuelnim predstavljanjem.

U cilju analize zašto su korišćeni određeni alati i algoritmi, radovi su detaljnije pregledani. Može se zaključiti da je fazi algoritam (engl. *Fuzzy miner*) najkorišćeniji algoritam, koji se skoro podjednako koristi u *ProM* (45%) i *Disco* (55%) alatima. Međutim, heuristički algoritam (engl. *Heuristic Miner*), Genetski algoritam i α -algoritam su korišćeni samo u *ProM*-u.

Štaviše, kada su istražene perspektive slučaja, vremena i organizacije, istraživači su koristili dodatne tehnike (obično *ProM* dodatke), kao što su tačkasti grafikon, analiza sekvence performansi, obojena Petrijeva mreža, dodatak za otkrivanje društvenih mreža i algoritam za primopredaju posla (engl. *handover of work*). Kada su vršili detaljniju analizu koristeći *Disco*, istraživači su koristili *fazi algoritam* i analizu varijanti toka procesa (engl. *trace variant analysis*). Iako nije tako bogat tehnikama kao *ProM*, *Disco* se pokazao kao jednostavan za korišćenje, veoma brz nad velikim dnevnicima događaja dok otkriva znanje visokog nivoa, sposoban da rukuje složenim procesima, posebno od pomoći za prethodnu obradu i izvoz podataka i generisanje intuitivnih mapa procesa koje se mogu uvećati i umanjiti da bi se obezbedilo više detalja. Štaviše, sa implementacijom fazi algoritma, *Disco* je bio u mogućnosti da istraži perspektive slučaja i vremena, iako su za detaljnije i pronicljivije analize istraživači preferirali *ProM*, s obzirom na njegove brojne dodatke.

Rezultati predstavljenog pregleda literature o primeni automatskog otkrivanja poslovnih procesa u industriji su pokazali da se perspektiva toka kontrole najviše koristi, prateći perspektivu slučaja i vremena i organizacione perspektive. Perspektive slučaja i vremena jasno pokazuju važnost i korisnost automatskog otkrivanja poslovnih procesa, jer je svaka primena perspektive slučaja i vremena rezultirala pozitivnim ishodom. S obzirom na tipove automatskog otkrivanja poslovnih procesa, prilikom provere usaglašenosti, istraživači su koristili *ProM* u čak 96% radova, što implicira da za ovu konkretnu aktivnost *ProM* nudi pogodnije tehnike nego *Disco*. Štaviše, nekoliko važnih zapažanja je dato u pogledu industrije. Slično primeni istraživanja i eksploatacije podataka u zdravstvenoj industriji, koja ima veliki potencijal da omogući zdravstvenim sistemima da sistematski koriste podatke i analitiku za identifikaciju neefikasnosti i najbolje prakse. Zdravstvena industrija se razlikovala od prosečnih vrednosti za sve industrije, sa malim procentom istraživača koji su se fokusirali na proveru usaglašenosti i sa širokom upotrebom *Disco* softverskog alata, umesto *ProM*-a, za otkrivanje toka kontrole procesa. IKT industrija je, međutim, pokazala natprosečnu upotrebu provere usaglašenosti, koristila je samo *ProM* za sve obavljene aktivnosti i nije koristila ni u jednom radu ni vremensku perspektivu, što je rezultiralo bez poboljšanja i predloga za poboljšanje. S obzirom na softverske alate, čini se da je *ProM* najdominantniji alat, iako je važno naglasiti da je *Disco* korišćen jednako kao i *ProM*, zbog svoje pogodnosti i praktičnosti, kada se posmatra isključivo otkrivanje modela procesa.

Rezultati predstavljenog pregleda literature o primeni automatskog otkrivanja poslovnih procesa u industriji su pokazali da se tipovi i perspektive ne primenjuju jednako u svakoj industriji, kao i da se

vrste analiziranih poslovnih procesa razlikuju. Samim tim, u različitim scenarijima su primenjeni različiti tipovi, algoritmi, tehnike i softverski alati. Kako bi se dalje ispitao odnos između ključnih faktora primene automatskog otkrivanja poslovnih procesa, sproveden je sistematski pregled literature samo u proizvodnoj industriji, gde su primenjene naprednije statističke metode kako bi se ispitala njihova povezanost. Naredno potpoglavlje predstavlja dobijene rezultate.

2.2.2. Pregled primene automatskog otkrivanja poslovnih procesa u proizvodnim industrijskim sistemima

Ovaj sistematski pregled literature je imao za cilj da sintetizuje znanja o primenama automatskog otkrivanja poslovnih procesa u proizvodnoj industriji, tako što će izdvojiti činjenične podatke iz studija slučaja koje izveštavaju o analizi stvarnih podataka. Motivacija za istraživanje te teme jeste detaljnija analiza pretpostavki o vezama između faktora koji potencijalno utiču na tok izvršavanja analize procesa primenom automatskog otkrivanja poslovnih procesa. Samim tim rezultati ovog sistematskog pregleda literature su korišćeni pri razvoju istraživačkog modela i hipoteza koji se odnose na IP1. Rezultati rada su objavljeni u međunarodnom časopisu [28] i na međunarodnoj konferenciji [27].

Ključni elementi istraživanja su ciljevi projekata automatskog otkrivanja poslovnih procesa, posmatrani poslovni procesi, informacioni sistemi i oprema koja podržava izvođenje poslovnih procesa, tipovi automatskog otkrivanja poslovnih procesa, algoritmi, tehnike i softverski alati. Podaci su prikupljeni putem sistematskog pregleda literature, koji se sastoji od planiranja, izvršenja i faze izveštavanja, nakon čega sledi deskriptivna statistika i unakrsna analiza. Rezultati analize podataka predstavljaju distribuciju i učestalost svakog ključnog elementa istraživanja zajedno sa otkrivenim odnosima među njima. Za svaki otkriveni cilj projekata automatskog otkrivanja poslovnih procesa, tehnika unakrsnih tabela predstavlja procese analizirane u okviru cilja, zajedno sa tipovima automatskog otkrivanja poslovnih procesa, algoritmima i tehnikama primenjenim u određenim scenarijima.

Definisana istraživačka pitanja su sledeća:

1. Koji su ciljevi postavljeni za projekte automatskog otkrivanja poslovnih procesa u proizvodnoj industriji?
2. Koji poslovni procesi se analiziraju i koji informacioni sistemi ih podržavaju?
3. Koje vrste i perspektive automatskog otkrivanja poslovnih procesa su izvedene i koji alati i tehnike su primenjeni?
4. Da li postoji veza između ciljeva projekta, analiziranih poslovnih procesa i primenjenih tehnika?

Strategija pretraživanja zasnovana na početnom opsegu odredila je indeksne baze podataka, termine za pretragu, kriterijume uključivanja i isključivanja, kao i potrebu za dodatnim ručnim pretraživanjem i pretraživanjem referenci unazad i unapred. Početne indeksne baze podataka koje su automatski pretražene su *Scopus* i *Web of Science*. Ključne reči sadržane u izrazu za pretragu su:

„*Process mining*“ AND *manufacturing* AND („*case study*“ OR *application*) AND *Year* ≥ 2009.

Da bi se obezbedio kvalitet prikupljenih podataka, kriterijumi za ocenjivanje primarnih studija su fokusirani na identifikaciji samo radova koji pružaju dovoljno informacija da bi se omogućila sinteza u smislu odgovora na postavljena istraživačka pitanja. Kriterijumi za izbor primarnih studija podeljeni su na sledeće kriterijume za uključivanje i isključivanje:

- Kriterijum isključenja 1: Duplikati radova koji se nalaze u više od jedne indeksne baze podataka treba da budu isključeni;
- Kriterijum isključenja 2: Radove koji pominju automatsko otkrivanje poslovnih procesa, ali primenjuju druge tehnike u studiji slučaja, treba ukloniti;
- Kriterijum za uključivanje 1: Rad treba da predstavi studiju slučaja primene automatskog otkrivanje poslovnih procesa u kontekstu proizvodnje.

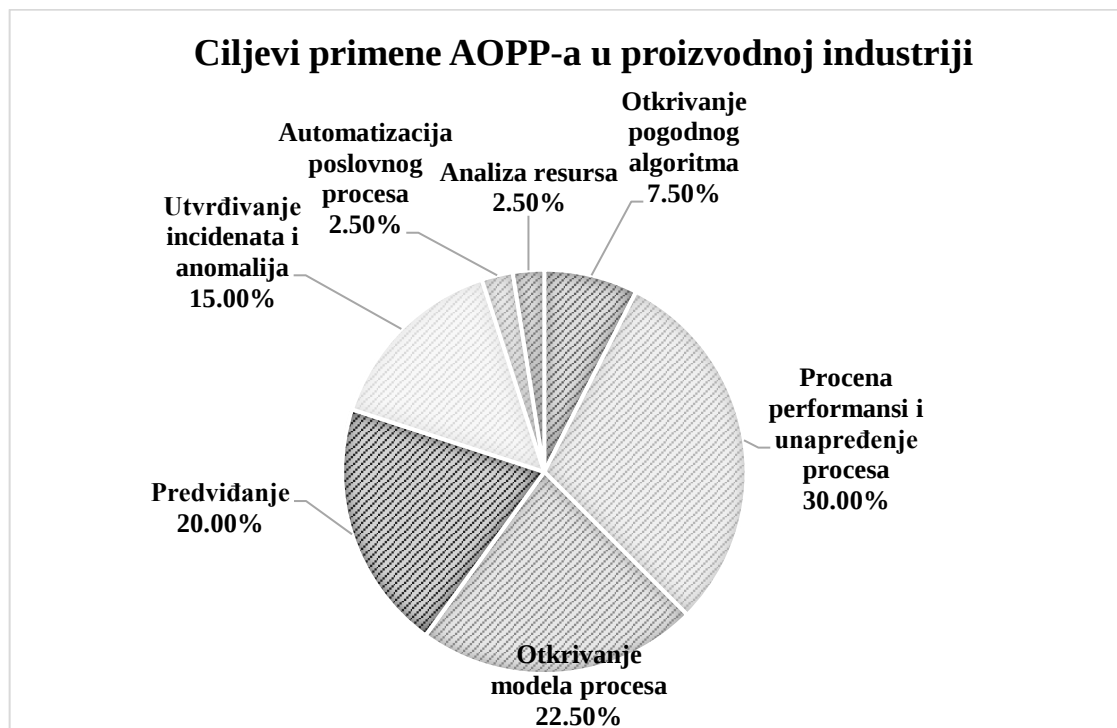
Nakon primene kriterijuma za izbor primarnih studija, opšti kriterijumi kvaliteta u vezi sa odabranim primarnim studijama su postavljeni da priznaju primarne studije iz opšteprihvaćenih

časopisa, mesta, radionica i beleški sa predavanja ili primarnih studija koje su predstavljale naučno utemeljen pristup. Kriterijumi kvaliteta nisu bili rigorozni, jer je cilj bio da se sintetizuje širi spektar prijavljenih studija slučaja kako bi se omogućila uspešna primena tehnika analize podataka. Konačno, strategija ekstrakcije podataka je razvijena na osnovu ključnih istraživačkih varijabli i ključnih istraživačkih pitanja.

Nakon primene izraza za pretragu za automatsko pretraživanje indeksnih baza podataka, početna pretraga je rezultirala sa 69 radova na *Scopusu* i 42 rada u *Web of Science* indeksnoj bazi. Izvršena je dodatna ručna pretraga radi pronalaženja radova istaknutih autora iz ove oblasti, dodajući još 7 radova koji sadrže definisani termin za pretragu u naslovu, sažetku ili ključnim rečima. Kriterijumi isključenja i uključivanja su zatim primenjeni na ovih 118 radova da bi se otkrile primarne studije za dalju analizu. Prvo je uklonjeno 20 duplikata. Na osnovu naslova i sažetka preostalih radova, zaključeno je da je 56 radova predstavljalo studiju slučaja analize procesa u proizvodnoj industriji. Od tih 56 radova, 19 primenjenih nije primenilo tehnike koje bi se mogle prepoznati kao specifične tehnike automatskog otkrivanja poslovnih procesa, već tehnike analize podataka kao što su mašinsko učenje ili statistika. Pretraživanje referenci unazad i unapred primenjeno je na preostalih 37 radova, otkrivajući još tri primarne studije koje se nisu pojavile kroz početnu automatsku pretragu. Konačan broj primarnih studija dobijenih za dalju analizu je 40. Izabrane primarne studije su uvezene u *Mendeley*, a prethodno definisane varijable su ekstrahovane tokom čitanja celog teksta.

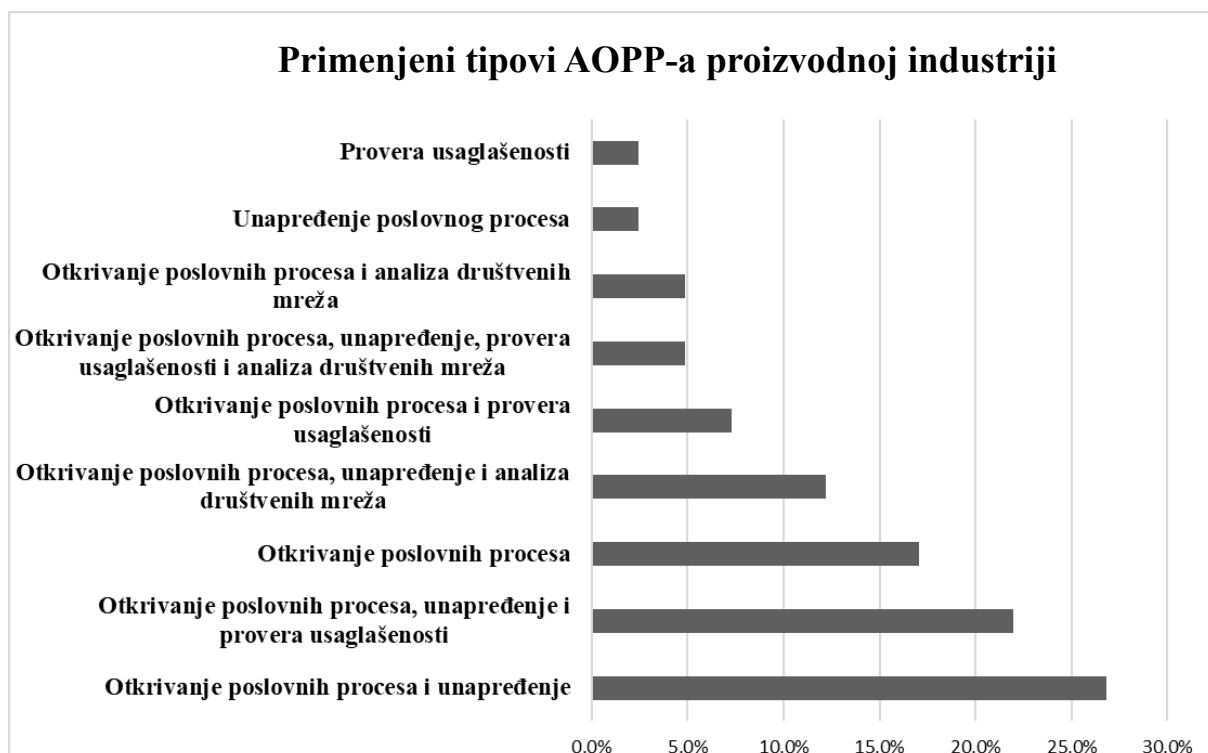
Posmatrane promenljive su kategoričkog tipa, te se na njih može primeniti deskriptivna statistika i neparametrijski testovi, poput unakrsne tabele (engl. *cross-tabulation*) u *IBM SPSS* statističkom alatu [68]. Izvučeni podaci nisu pogodni za druge statističke tehnike na skupovima podataka sa kategoričkim varijablama kao što su logistička regresija i hi-kvadrat test, pošto je uzorak premali da bi rezultati bili pouzdani. Opcija deskriptivne statistike i učestalosti iz *IBM SPSS*-a bila primenjena na sve varijable kako bi se stekao uvid u njihovu distribuciju i odgovorili na prva tri istraživačka pitanja navedena u proceduri sistematskog pregleda literature. Istraživačko pitanje broj 4 je ključno pitanje za doprinos ovog sistematskog pregleda literature. Da bi se otkrila mogućnost veze između varijabli, primenjena je tehnika unakrsnih tabela.

Izdvojeni ciljevi koji su postavljeni u okviru projekata analize poslovnih procesa u proizvodnoj industriji su prikazani na slici 7. Otkrivanje modela poslovnih procesa, evaluacija performansi i unapređenje procesa, kao i predviđanje zauzimaju 72.5% svih ciljeva projekta. Slede otkrivanje incidenata i anomalija [69–73], otkrivanje najboljih algoritama [74–76], automatizacija poslovnih procesa [77] i analiza resursa [78]. Cilj otkrivanja modela procesa je postavljen za studije slučaja u kojima je fokus bio na primeni algoritama otkrivanja procesa radi pronalaženja kontrolnog toka procesa [50,79–84]. Evaluacija performansi i unapređenje procesa je širi cilj, i u posmatranim studijama slučaja uključuje analizu uzroka problema unutar procesa (engl. *root-cause analysis*) i nestabilnosti prilikom planiranja proizvodnje [56], analizu performansi procesa [85–91], analizu društvenih mreža i analiza varijanti putanja procesa [92]. Može se zaključiti da je, pored opštih ciljeva kao što su otkrivanje i unapređenje procesa, u posmatranim studijama slučaja veoma važno predviđanje. Primarne studije čiji je cilj bio korišćenje automatskog otkrivanja poslovnih procesa radi predviđanja unutar procesa su uspešno predvidele troškove procesa izvršenja proizvodnje [93], intervale inspekcije održavanja, greške u samom proizvodnom procesu [94–96], vreme izvršenja proizvodnje [97], način izvršavanja procesa pre implementacije sistema za podršku planiranja resursa za proizvodnju (engl. *Enterprise Resource Planning – ERP*) [98], ponašanje tehnološkog sistema [99] i opterećenje resursa [100].



Slika 7. Ciljevi primene automatskog otkrivanja poslovnih procesa u proizvodnoj industriji

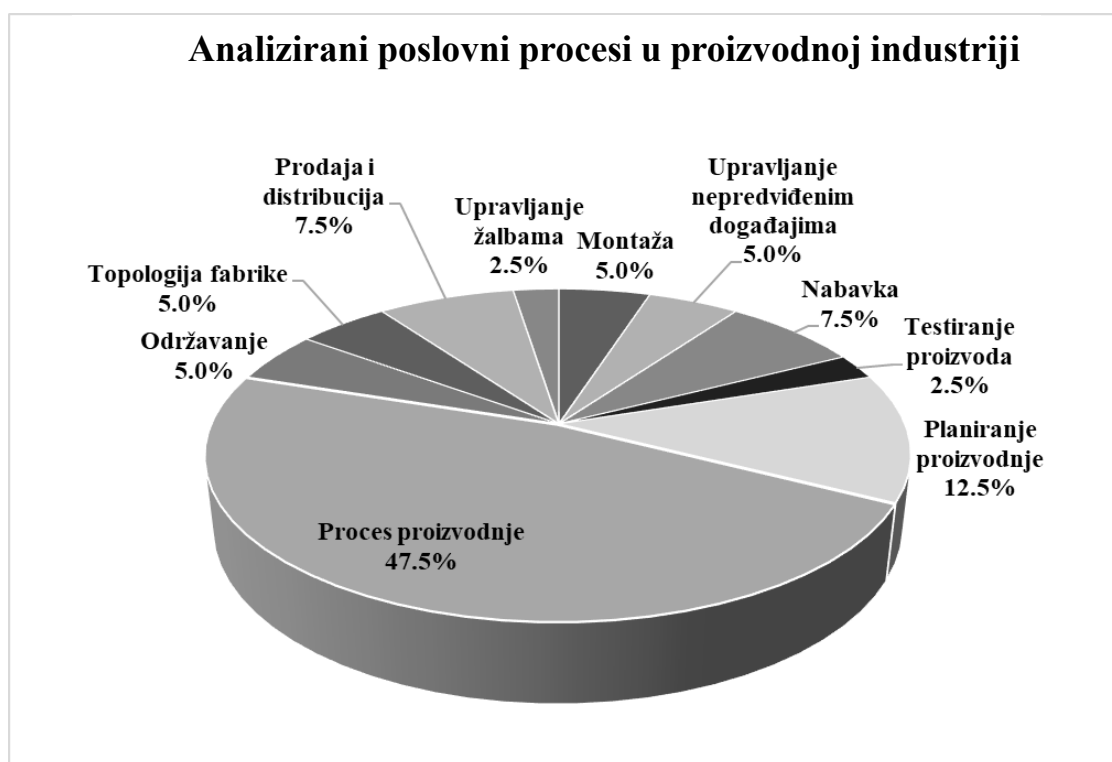
Na slici 8 prikazani su tipovi i perspektive procesnog rudarenja, koje se obično primenjuju kombinovano, u zavisnosti od cilja analize. Kada se posmatra odvojeno, tip otkrivanja procesa je primenjen u 95.1% studija slučaja, poboljšanje procesa u 68.3%, provera usaglašenosti u 34.1% i analiza društvenih mreža u 24.4%.



Slika 8. Kombinacije primene tipova automatskog otkrivanja poslovnih procesa u proizvodnoj industriji

Slika 9 prikazuje raspodelu analiziranih poslovnih procesa u posmatranim primarnim studijama. Može se zaključiti da je najznačajni poslovni proces zapravo sam proces proizvodnje proizvoda, koji je bio predmet analize u 47.5 % posmatranih studija slučaja. Nakon njega slede procesi planiranja

proizvodnje, nabavke i prodaje i distribucije, dok su ostali poslovni procesi zastupljeni u 5% ili manje posmatranih studija slučaja.



Slika 9. Distribucija poslovnih procesa proizvodne industrije analiziranih u posmatranim studijama slučaja

Prilikom analize algoritama i tehnika primenjenih u studijama slučaja u proizvodnoj industriji, prvo su posmatrani algoritmi za otkrivanje poslovnih procesa. Fazi algoritam (engl. *fuzzy miner*) se primenjuje u 35.1% posmatranih studija slučaja, uglavnom u softverskom alatu *Disco* [73,86–88,94,95,97,100]. U nekim slučajevima je primenjen u novom alatu ili prototipu alata razvijenom od strane autora studije slučaja [91,101], kao i u softverskim alatima *Celonis* (Sanchez et al., 2020) i *ProM* [88]. Sledeći često primenjivani algoritam za otkrivanje poslovnih procesa je heuristički algoritam, primenjen u 29.7% posmatranih studija slučaja. Heuristički algoritam se uglavnom primenjuje kao dodatak u alatu *ProM* [41,50,69,74,75,78,82,85], iako je u nekim slučajevima primenjen u novom alatu ili prototipu [84]. Induktivni algoritam je treći istaknuti algoritam za otkrivanje poslovnih procesa, jer se primenjuje u 24.3% posmatranih studija slučaja. U svim posmatranim slučajevima, induktivni algoritam se primenjuje u alatu *ProM* kao dodatak [69,70,80,85,92,99,102,103]. Drugi algoritmi otkrivanja poslovnih procesa se primenjuju ređe u studijama slučajeva sa specifičnim ciljevima.

Primenjene tehnike za unapređenje poslovnih procesa bile su različite, iako je u većini posmatranih studija slučaja primenjeno izračunavanje vremena trajanja procesa i analiza uskih grla, jer su to osnovne tehnike analize performansi. Proračun vremena trajanja procesa primenjen je u 75% posmatranih slučajeva, a analiza uskog grla u 71.4% posmatranih slučajeva. Ove dve tehnike su obično primenjivane zajedno da bi se stekao uvid u vremensku perspektivu analize procesa i otkrile aktivnosti koje izazivaju kašnjenja.

Disco i *ProM* su podjednako podržavali izračunavanje vremena protoka i analizu uskih grla. Otkrivanje povratnih petlji (engl. *feedback loops*) u procesu je obavljena u 21.4% posmatranih slučajeva u kojima je vršen tip unapređenja poslovnih procesa, podjednako u alatima *ProM* i *Disco* [56,86,88,101]. Analiza varijanti izvršenja procesa primenjena je u 17.9% posmatranih studija slučaja koje su izvršile tip unapređenja procesa [56,92,95,98], dok je tačkasta analiza grafikona (engl. *Dotted-chart analysis*) u alatu *ProM* primenjena u 14.3% posmatranih studija slučaja koje su izvršile unapređenja procesa [71,86,104]. Specifična tehnika iz *ProM*-a, dodatak za analizu performansi, primenjena je u 10.7%

posmatranih studija slučaja da bi se izračunale mere performansi, kao što su vreme čekanja i vreme trajanja aktivnosti [74,79,86]. Dodatne tehnike unapređenja primenjene u manje od 10% posmatranih studija slučaja koje su izvršile poboljšanje procesa.

Da bi izvršili organizacionu perspektivu, praktičari su se uglavnom oslanjali (37.5%) na dodatak za otkrivanje društvenih mreža iz *ProM*-a za otkrivanje iskorišćenosti tehnoloških sistema i odnosa između radnih zadataka i tehnoloških sistema [79], proračun opterećenja tehnoloških sistema [75], kao i za dijagram primopredaje zadataka [86]. Neki praktičari [84] su razvili novi alat za automatsko otkrivanje poslovnih procesa, sa analizom uloga resursa zasnovanom na radnoj metrici slično onom dostupnom u *ProM*-u, dok su autori u [95] posmatrali efikasnost resursa u alatu *Disco*. Ostale primene su bile otkrivanje procesa korišćenja mašine sa induktivnim algoritmom u alatu *ProM* [99] i razvoja društvene mreže utrošenih resursa [76]. S obzirom na softverske alate, *ProM* i *Disco* su se obično primenjivali zajedno. Ako se njihova apsolutna učestalost primene posmatra odvojeno, *ProM* se primenjuje u 56.4% i *Disco* u 33.3% posmatranih studija slučaja.

Kako bi se odgovorilo na istraživačko pitanje 4 ovog sistematskog pregleda literature, kreirana je tabela kontingencije u IBM SPSS alatu, kako bi se utvrdili paterni prilikom primene automatskog otkrivanja poslovnih procesa u proizvodnoj industriji. Rezultati analize su detaljno predstavljani i obrazloženi u [28], a u narednom tekstu će biti predstavljen njihov značaj za dalje istraživanje autora i doktorsku disertaciju.

Doprinos ovog istraživanja može se posmatrati kroz teorijske i praktične implikacije. Sistematski pregled literature studija slučaja koje predstavljaju aplikacije automatskog otkrivanja poslovnih procesa u domenu proizvodnje značajan je za praktičare automatskog otkrivanja poslovnih procesa i zainteresovane strane u proizvodnim sistemima jer pruža uvid u najsavremenije, najčešće probleme i vrste procesa od interesa, kao i analizu primenjenih pristupa i ishoda. Pošto su poslovni analitičari i menadžeri u proizvodnim organizacijama uglavnom radili u proizvodnoj industriji, oni nemaju potrebne osnovne informacije i iskustvo u upravljanju i izvođenju projekata automatskog otkrivanja poslovnih procesa. Štaviše, kako se automatsko otkrivanje poslovnih procesa češće primenjuje u domenu zdravstvene nege i obrazovanja nego u proizvodnji [86,101], stručnjaci za automatsko otkrivanje poslovnih procesa možda nemaju dovoljno iskustva u ovoj industriji. Ovo istraživanje predstavlja pregled i trendove u primeni metoda i može biti dragoceno sredstvo za fazu planiranja projekata automatskog otkrivanja poslovnih procesa koji se izvode u proizvodnim organizacijama, kada treba odrediti cilj projekta, tipove automatskog otkrivanja poslovnih procesa i odgovarajuće algoritme/tehnike.

Ciljevi postavljeni za projekte automatskog otkrivanja poslovnih procesa pokazali su da rezultati analize mogu imati praktične implikacije u donošenju odluka i redizajnu procesa u proizvodnim sistemima. Tehnike automatskog otkrivanja poslovnih procesa bile su pogodne za otkrivanje modela procesa, i još važnije, za procenu performansi procesa. Performanse procesa su uglavnom procenjene izračunavanjem ukupnog vremena protoka izvršenja procesa, otkrivanjem uskih grla i povezanim vremenima čekanja. Štaviše, otkrivanje incidenata i anomalija u izvršavanju procesa omogućilo je organizacijama da prilagode intervale pregleda održavanja uređaja i opreme. Inicijalna analiza varijabli istraživanja je takođe pokazala da su pored procesa izvođenja proizvodnje u fokusu proizvodnih sistema bili i osnovni procesi kao što su planiranje proizvodnje, prodaja, distribucija i nabavka. Ubedljivo najprimenjeniji informacioni sistem u proizvodnim sistemima bio je ERP, a zatim Sistem za upravljanje izvršenjem proizvodnje (engl. *Manufacturing Execution System* – MES). Podaci o događajima su takođe često ekstrahovani iz proizvodne opreme i uređaja.

Takođe, značajna zapažanja mogu se dati u vezi sa specifičnim algoritmima, tehnikama i softverskim alatima analiziranim u ovom pregledu literature. Poslovni analitičari su uglavnom imali za cilj da otkriju kontrolne tokove procesa, procene performanse procesa i provere kvalitet otkrivenog modela. Stoga su otkrivanje, poboljšanje i usklađenost često primenjivani istovremeno. Proizvodni sistemi ne odstupaju od ostalih domena u korišćenju algoritama za otkrivanje, jer su fazi algoritam, heuristički algoritam i induktivni algoritam pristupi koji trenutno proizvode najtačnije i najispravnije modele procesa u realnim scenarijima [86]. Zanimljivo je otkriti da se fazi algoritam, koji se primenjuje

u većini studija slučaja, ne koristi često za otkrivanje modela procesa izvršenja proizvodnje. Fazi algoritam se obično primenjuje na složene i nestrukturirane evidencije događaja, što znači da su evidencije događaja izvučene iz uređaja i opreme koja podržava proizvodno izvršenje dovoljno strukturirane da heuristički i induktivni algoritam proizvode modele procesa visokog kvaliteta. Štaviše, istaknute tehnike poboljšanja koje su široko primenjene u analiziranim studijama slučaja bile su analiza uskih grla i izračunavanje vremena protoka. S druge strane, tehnike analize društvenih mreža su primenjivane obilno i uglavnom u odnosa između proizvodnih resursa.

Takođe, kreirana tabela kontigencije, odnosno unakrsna analiza ciljeva, poslovnih procesa, tipova procesnog rudarenja i tehnika omogućila je uvid u obrasce primene automatskog otkrivanja poslovnih procesa u proizvodnim sistemima, motivišući autora da ispita odnose između ključnih faktora koji utiču na primenu automatskog otkrivanja poslovnih procesa.

Zaključeno je da, na posmatranom uzorku, postoji potencijalna veza između ciljeva projekata i sprovedenih tipova i perspektiva AOPP-a. Ostale pretpostavke su se pokazale kao netačne, na ovom uzorku. Razlog za ovakve rezultate može biti specifičnost proizvodne industrije, mali uzorak, nepotpuni podaci, kao i sličnost posmatranih studija slučaja. Ovo istraživanje je doprinelo definisanju ključnih činioca istraživanja doktorske disertacije, kao i definisanju istraživačkih modela i hipoteza. Takođe, zaključeno je da u studijama slučaja nema dovoljno podataka koji ukazuju na kvalitet dnevnika izvršenja događaja.

2.2.3. Pregled problema kvaliteta podatka i tehnika čišćenja dnevnika izvršenja događaja

Naredna potpoglavlja predstavljaju sintezu problema kvaliteta podataka dnevnika izvršenja događaja nastalu pregledom relevantne literature i rezultate pregleda literature o primeni tehnika čišćenja dnevnika izvršenja događaja.

2.2.3.1. Sinteza problema kvaliteta podataka dnevnika izvršenja događaja

Kao što je spomenuto u uvodu, Bose [7] i Suriadi [8] su definisali četiri široke kategorije problema kvaliteta podataka (nedostajući podaci, netačni podaci, neprecizni podaci i irelevantni podaci), sa definisanim obrascima imperfekcije (OI) koji opisuju specifične manifestacije problema kvaliteta podataka. U tabeli 3 autor je spojio probleme kvaliteta dnevnika i obrasce imperfekcije kako bi se stekao uvid u prominentne probleme i njihovu manifestaciju.

Tabela 3. Manifestacija problema kvaliteta podataka i obrazaca imperfekcije

Entiteti dnevnika izvršenja događaja								
		Slučaj	Događaj	Veza	Atribut slučaja/događaj a	Pozicija/ Vremenska odrednica	Naziv aktivnosti	Resurs
Problemi kvaliteta podataka	Nedostajući podaci	Pr.1	Pr.2, OI1	Pr.3, OI2	Pr.4, Pr.9	Pr.5, Pr.7	Pr.6	Pr.8
	Netačni podaci	Pr.10	Pr.11	Pr.12, OI3	Pr.13, Pr.18	Pr.14, Pr.16, OI6, OI7, OI8	Pr.15, OI4, OI5	Pr.17, OI4
	Neprecizni podaci	/	/	Pr.19	Pr.20, Pr.25, OI9	Pr.21, Pr.23, OI8	Pr.22, OI10	Pr.24
	Irelevantni podaci	Pr.26	Pr.27, OI6, OI11	/	/	/	/	/

OI1 – Raštrkani događaj; OI2 – Neuhvatljiv slučaj; OI3 – Raštrkani slučaj; OI4 – Zagađeni naziv atributa; OI5 – Iskrivljeni naziv atributa; OI6 – Hvatanje događaja zasnovano na obrascu; OI7 – Nehotično putovanje kroz vreme;

OI8 – Neukorenjeni događaj; OI9 – Sinonimni naziv atributa; OI10 – Istoimeni naziv atributa; OI11 – Kolateralni događaji.

Kategorije problema kvaliteta podatka mogu se definisati na sledeći način.

Kategorija nedostajućih podataka odnosi se na problem kvaliteta gde podatak nedostaje u dnevniku izvršenja događaja, iako je obavezan. Na primer, „Problem nedostajućih podataka slučaja se odnosi na scenario gde je slučaj izvršen u stvarnosti, ali nije zabeležen u dnevniku izvršenja događaja“ [7]. Kategorija netačnih podataka odnosi se na problem kvaliteta gde su podaci dati, ali su pogrešno evidentirani. Na primer, „Problem netačnih slučajeva odgovara scenariju gde određeni slučajevi u dnevniku izvršenja događaja pripadaju drugom procesu“ [7]. Kategorija nepreciznih podataka odnosi se na probleme kvaliteta gde su evidentirani unosi previše uopšteni, što dovodi do gubitka preciznosti. Na primer, „Problem sa nepreciznim nazivima aktivnosti odgovara scenariju gde unutar dnevnika izvršenja događaja može postojati više različitih događaja sa istim imenom aktivnosti“ [7]. Kategorija irelevantnih podataka odnosi se na pitanja kvaliteta gde evidentirani unosi mogu biti beznačajni za analizu. Na primer, „Problem irelevantnih slučajeva odgovara scenariju gde se određeni slučajevi u dnevniku izvršenja događaja događaja smatraju nebitnim za određeni kontekst analize“ [7].

Prethodno opisani problemi kvaliteta se mogu manifestovati kroz različite entitete dnevnika izvršenja događaja [7]:

- Entitet slučaja odgovara instanci procesa koja je izvršena;
- Entitet događaja odgovara aktivnosti unutar procesa;
- Entitet veze odgovara asocijaciji između slučaja i događaja;
- Entitet atributa slučaja i događaja se odnosi na dodatne informacije koje slučaj ili događaj mogu imati;
- Entiteti pozicije i vremenske odrednice predstavljaju zabeleženo vreme izvršavanja aktivnosti, gde pozicija opisuje poziciju događaja unat dnevka izvršenja događaja, a vremenska odrednica opisuje stvarno vreme izvršenja događaja.;
- Entitet naziva aktivnosti se odnosi na naziv zabeleženih aktivnosti;
- Entitet resursa se odnosi na resurse utrošene na izvođenje aktivnosti.

Iz tabele 3, može se videti da informacije u vezi sa svim entitetima dnevnika izvršenja događaja mogu nedostajati, uključujući nedostajući slučaj, događaj, vezu, atribut slučaja/događaja, poziciju i vremensku odrednicu, naziv aktivnosti i resurs. Kada se posmatraju obrasci imperfekcije u vezi sa podacima koji nedostaju, ukoliko podaci o događaju nedostaju, javlja se obrazac raštrkanog događaja (OI1). Obrazac raštrkanog događaja opisuje scenario gde jedan zabeleženi događaj sadrži izostavljene informacije o drugim događajima koji su se desili tokom izvršavanja procesa. Još jedan obrazac imperfekcije koji razmatra podatke koji nedostaju je obrazac neuhvatljivog slučaja (OI2), koji se odnosi na scenario u kojem nedostaju informacije o odnosu između događaja i slučajeva, što uzrokuje problem da neki događaji nisu povezani sa određenim identifikatorom slučaja.

Problem netačnih podataka se takođe može manifestovati kroz sve entitete dnevnika događaja, sa mnogim obrascima imperfekcije dnevnika događaja. Obrazac imperfekcije raštrkanog slučaja (OI3) se javlja kada se odnos između događaja i slučaja evidentira pogrešno, jer neki događaji nedostaju u dnevniku izvršenja događaja koji se analizira. Netačne vremenske odrednice su uobičajen problem kvaliteta podataka, što dovodi do tri različita obrasca imperfekcije dnevnika izvršenja događaja. Hvatanje događaja zasnovano na obrascu (OI6) opisuje scenario kada se podaci evidentiraju iz direktno iz stranice neke aplikacije, što rezultira da su svi snimljeni podaci koje stranica sadrži sačuvani u isto vreme i stoga imaju istu vremensku oznaku. Obrazac nehotičnog putovanja kroz vreme (OI7) se javlja zbog korisničke greške, kada korisnik sistema slučajno unese netačnu vremensku odrednicu zbog njene blizine vremenskoj odrednici prethodno izvršenog događaja. Obrazac neukorenjenog događaja (OI8) nastaje kada je format datuma vremenske odrednice netačan. Entitet naziva aktivnosti često sadrži netačne podatke, što dovodi do zagađenih i iskrivljenih naziva. Obrazac zagađenog naziva atributa (OI4) opisuje situaciju kada su neki atributi događaja (kao što su naziv aktivnosti i resurs) strukturno isti, ali

se međusobno razlikuju po svojim stvarnim vrednostima. Obrazac iskrivljenog naziva (OI5) se javlja kada su dva naziva atributa sintaksički i semantički slična, ali nisu zabeležena kao potpuno iste vrednosti.

Problem sa nepreciznim kvalitetom podataka može se otkriti unutar odnosa između slučaja i događaja, atributa slučaja/događaja, pozicije/vremenske odrednice, naziva aktivnosti i resursa. Prethodno pomenuti obrazac neukorenjenog događaja (OI8) koji opisuje netačnu vremensku odrednicu, takođe se može kategorisati u okviru problema sa nepreciznom vremenskom odrednicom. Dodatni obrasci imperfekcije u vezi sa nepreciznim podacima su obrazac sinonimnih naziva (OI9), koji se javlja kada ista aktivnost ima različite nazive aktivnosti u okviru jednog dnevnika događaja, i obrazac homonimnih naziva (OI10), koji se javlja kada se različite aktivnosti istog procesa evidentiraju sa istim nazivom aktivnosti.

Irelevantni podaci u vezi sa informacijama o slučajevima i događajima se takođe mogu naći u dnevniku izvršenja događaja. Prethodno pomenuti obrazac hvatanja događaja zasnovano na obrascu (OI6) dovodi i do problema evidentiranja irrelevantnih događaja. Obrazac imperfekcije specifičan samo za irrelevantne podatke o događajima je obrazac kolateralnih događaja (OI11), koji opisuje situaciju kada brojni događaji opisuju isti korak u procesu.

Pored navedenih problema kvaliteta podataka, značajno je spomenuti i probleme kvaliteta koji su uzrokovani od strane procesnih karakteristika, a to su [7]:

- Velike količine podataka, odnosno velik broj zabeleženih slučajeva i događaja procesa;
- Kompleksnost podataka, odnosno velik broj različitih putanja izvršavanja procesa;
- Nivo detaljnosti događaja, odnosno velik broj različitih aktivnosti.

Predstavljeni problemi kvaliteta dnevnika izvršenja događaja su bitne dimenzije istraživanja koje se tiču IP2, što će dalje u radu biti obrazloženo.

2.2.3.2. Sinteza tehnika čišćenja dnevnika izvršenja događaja

Kako bi se sintetizovali podaci o primeni tehnika čišćenja dnevnika izvršenja događaja, sproveden je sistematski pregled literature. S obzirom da ovaj pregled literature nije glavni doprinos rada, biće predstavljen sažeto, a ceo postupak je objavljen u okviru rada u časopisu [29]. Postoji jedan prethodni sistematski pregled literature koji je objavljen na temu tehnika čišćenja dnevnika izvršenja događaja [14], koji je referenciran kao smernica za grupisanje tehnika.

Cilj pregleda literature je bio otkrivanje najčešće primenjivanih tehnika za rešavanje problema kvaliteta podataka u oblasti automatskog otkrivanja poslovnih procesa, kao i njihovo grupisanje u dimenzije istraživanja. Stoga, istraživačko pitanje kojim se pregled literature vodio je:

Koje su najčešće tehnike manipulacije podacima primenjene radi rešavanja problema kvaliteta podataka?

Strategija pretrage relevantnih studija je bila sačinjena od termina pretrage, baza podataka, kriterijuma uključivanja i isključivanja radova i strategije prikupljanja podataka. Definisani termin za pretragu je sledeći:

(TITLE-ABS-KEY („event log“) AND (TITLE-ABS-KEY („data quality“) OR TITLE-ABS-KEY („pre*processing“) OR TITLE-ABS-KEY (cleaning) OR TITLE-ABS-KEY (filtering) OR TITLE-ABS-KEY (repairing))).

Pretražene indeksne baze su bile *Web of Science*, *IEEE Xplore*, i *EBSCO*. Radovi sun morali biti napisani na engleskom jeziku i mogli su imati bilo koju godinu objavljivanja. Dalji kriterijumi za prihvatanje radova su bili sledeći:

- Kriterijum uključivanja 1: Rad mora da predstavlja tehniku ili pristup primenjen na rešavanje problema kvaliteta podataka.
- Kriterijum uključivanja 2: Rad mora da pripada oblasti automatskog otkrivanja poslovnih procesa. Slične discipline, poput mašinskog učenja i istraživanje i eksploatacije podataka ne mogu biti uključene.

- Kriterijum uključivanja 3: Rad mora da bude objavljen na konferenciji ili u časopisu.
- Kriterijum isključivanja 1: radovi objavljeni pre 2019 godine koji nisu citirani biće izbačeni.
- Kriterijum isključivanja 2: Radovi koji su analizirali probleme kvaliteta nad sintetičkim podacima će biti isključeni.

Strategija ekstrakcije podataka je postavljena da prikuplja podatke o problemima kvaliteta podataka, tehnikama čišćenja dnevnika izvršenja događaja i softverskim alatima unutar kojih je tehnika primenjena. Strategija pretrage je rezultovala sa 29 primarnih studija objavljenih između 2013 i 2023 godine, se većinom studija objavljenih u 2022 godini. Primarne studije su uglavnom objavljene na konferencijama (77% od svih primarnih studija), sa manjim učešćem radova iz časopisa (23% od svih primarnih studija).

Tehnike čišćenja dnevnika izvršenja događaja uočene u dnevnicima izvršenja događaja nastalim na osnovu stvarnih poslovnih procesa su grupisane u 8 kategorija na osnovu pristupa koji tehnika koristi kako bi umanjila ili uklonila problem kvaliteta podataka i na osnovu prethodno spomenute taksonomije tehnika [14]. Tabela 4 predstavlja kategorije tehnika sa pripadajućim tehnikama i softverskim alatima unutar kojih su primenjene. Svaka tehnika ima iskazanu učestalost primene u okviru svih analiziranih primarnih studija i svaka kategorija ima izraženu ukupnu učestalost primene. Većina primarnih studija nije navela informacije o softverskim alatima. Najčešće korišćeni softverski alati za čišćenje dnevnika izvršenja događaja su, na osnovu dostupnih informacija, *ProM*, *RapidProm*, *Matlab* i *CPNtools*.

Prema literaturi, najčešće primenjivana grupe tehnika za rešavanje problema kvaliteta podataka su grupa veštačka inteligencija, mašinsko učenje i duboko učenje, koja sadrži algoritme i tehnike primene u različitim scenarijima. Bajazove mreže (engl. *Bayesian networks*) su grupa probabilističkih grafičkih modela koji se mogu primeniti kako bi popravili delove dnevnika izvršenja događaja sa nedostajućim vremenskim odrednicama [105] i nedostajućim događajima [106,107]. Dodatno, LSTM (engl. *Long Short-Term Memory*) veštačke neuronske mreže iz oblasti dubokog učenja imaju mogućnost predviđanja nedostajućih događaja i naziva aktivnosti [108]. Još jedna tehnologija koja ima mogućnost da reši problem nedostajućih podataka su algoritmi zasnovani na verovatnoći, odnosno *Single Imputation by Event Relationship* (SIER) i *Multiple Imputation by Event Chain* (MIEC) algoritmi, koji imaju sposobnost popravke dnevnika sa nedostajućim događajima, vremenskim odrednicama i resursima [109]. Takođe, algoritam nasumične šume (engl. *random forest algorithm*) je algoritam klasifikacije iz oblasti mašinskog učenja koji imaju mogućnost uočavanja događaja sa netačnim vremenskim odrednicama [110]. Još jedan klasifikacioni algoritam je algoritam klasifikacije i regresije (engl. *Classification and Regression Tree* – CART) primenjen radi otkrivanja tendencije nedostajućih podataka bez popravke dnevnika izvršenja događaja [111]. Prirodna obrada jezika (engl. *Natural Language Processing* – NLP) je podoblast mašinskog učenja usko povezana sa veštačkom inteligencijom koja omogućava mašinama da razumeju ljudski jezik. U oblasti automatskog otkrivanja poslovnih procesa koristi se radi detekcije, preimenovanja ili uklanjanja nepreciznih događaja i naziva aktivnosti [112,113].

Klasterovanje je takođe tehnika mašinskog učenja koja analizira obrasce i veze unutar seta podataka kako bi identifikovala slične grupe ili klustere. U oblasti automatskog otkrivanja poslovnih procesa kao posebna grupa tehnika za čišćenje dnevnika izvršenja događaja izdvaja se klasterovanje putanja (engl. *trace clustering*). Putanja je varijanta izvršavanja procesa, gde jedan dnevnik izvršavanja može imati mnogo različitih putanja. Klasterovanje putanja je veoma efektivno u izolovanju putanja koje sadrže neuobičajeno ponašanje i za detekciju obrazaca imperfekcije. Najčešće se koristi za umanjene veličine, kompleksnosti i nivoa detaljnosti događaja dnevnika izvršenja događaja [114–116]. Neki istraživači su koristili klasterovanje putanja da otkriju sličnosti između putanja sa nekompletnih podacima i radi predviđanja nedostajućih aktivnosti na osnovu matrice sledljivosti događaja [117]. Takođe, klasterovanje putanja je nekada neophodna tehnika koja prethodi drugim kompleksnijim tehnikama poput analize zasnovane na statističkoj inferentnosti (engl. *Statistical inference-based clustering*) [118]. Klasterovanje na osnovu minimalnog raspona drveta (engl. *Minimum-Spanning Tree* – MST) ima mogućnost da detektuje neprecizne nazive aktivnosti, atributa događaja i informacije o resursima [119].

Tehnike filtriranja putanja/događaja spadaju u tehnike transformacije podataka, jer određuju verovatnoću pojave događaja ili putanja i uklanjaju događaje ili tragove sa manjom učestalošću pojavljivanja [14]. Filtriranje je jedan od osnovnih koraka čišćenja dnevnika izvršenja događaja, gde istraživači moraju da postave graničnu vrednost frekvencije koju putanja ili događaj moraju da ispune kako bi bili uključeni u prethodno obrađeni dnevnik događaja [116,120–122]. Štaviše, utvrđeno je da je algoritam grananja i spajanja za matematičku optimizaciju (engl. *branch and bound algorithm*) uspešan u filtriranju neuobičajenog ponašanja identifikacijom najpogodnijeg modela koji odgovara zabeleženom dnevniku događaja [123–125]. Druga specifična tehnika za filtriranje dnevnika događaja je dodatak za popravljjanje dnevnika izvršenja događaja (engl. *Repair log plug-in*), implementiran u softverskim alatima *ProM* i *RapidProM* [126,127]. Dodatak za popravljjanje dnevnika izvršenja događaja filtrira evidenciju događaja posmatrajući uslovne verovatnoće između nizova aktivnosti i uklanjanje neprecizne i netačne događaje i nazive aktivnosti.

Tehnike za popravku dnevnika izvršenja događaja predstavljaju posebnu grupu tehnika, jer imaju mogućnost popravke delova dnevnika događaja bez primene domenskog znanja. Tehnike koje imaju sposobnost da poprave sve putanje procesa pre otkrivanja modela procesa su heurističke tehnike za popravku dnevnika [128,129]. Heurističke tehnike za popravku dnevnika događaja identifikuju povratne strukture i česte sekvence događaja koje su odgovarajućeg kvaliteta, a preostale putanje procesa izdele u više putanja koje zatim popravljaju na osnovu prethodno definisanih putanja visokog kvaliteta. Potrebno je naglasiti da su neki od prethodno spomenutih pristupa iz grupe veštačke inteligencije, mašinskog učenja i dubokog učenja u mogućnosti da delimično poprave dnevnik izvršenja događaja. Ipak, ove tehnike koriste dodatne referentne modele procesa, npr. model procesa definisan na osnovu prethodnog domenskog znanja [106–108].

Grupa tehnika ugrađene obrade podataka predstavlja algoritme za otkrivanje poslovnih procesa koji imaju mehanizme za implicitno filtriranje neuobičajenih aktivnosti. Praktičari automatskog otkrivanja poslovnih procesa unapred konfigurišu parametre za filtriranje i preciznost modela. Primeri ovih algoritama su induktivni algoritam i *Integer Linear Programming* – ILP algoritam [130].

Tehnike poravnavanja u oblasti automatskog otkrivanja poslovnih procesa se odnose na set metoda za poređenje posmatrani dnevnik izvršenja događaja sa otkrivenim modelom procesa ili referentnim modelom. Spomenute tehnike imaju za cilj da otkriju nivo usaglašenosti između posmatranog ponašanja u dnevniku i očekivanog ponašanja definisanog referentnog modela ili procenu kvaliteta otkrivenog modela procesa na osnovu njegove usaglašenosti sa dnevnikom događaja. Provera usaglašenosti zasnovana na poravnanju može da se koristi za pronalaženje problema redosleda izvršenja događaja [105], [131]. Još jedan pristup zasnovan na proveru usaglašenosti je poravnanje zasnovano na trošku (engl. *cost-based alignment*) [107]. Poravnanje zasnovano na trošku podrazumeva dodeljivanje troška različitim vrstama devijacija između posmatranih događaja i modela procesa poput nedostajućih aktivnosti ili retkih aktivnosti.

Tehnike apstrakcije događaja se mogu primeniti kako bi se pojednostavili i generalizovali podaci događaja zabeleženi u dnevniku izvršenja događaja transformisanjem podataka o događajima niskog nivoa apstrakcije u reprezentacije višeg nivoa apstrakcije, što olakšava analizu značajnih zaključaka iz podataka. Pregledom literature pronađena je jedna primarna studija koja primenjuje semantičku apstrakciju kako bi se minimizirao obim, složenost i nivo detaljnosti događaja modela deklarativnih procesa [132].

Kategorija tehnika pod nazivom „drugo”, navedena u tabeli 4 sadrži tehnike koje nisu striktno tehnike obrade podataka, ali su korišćene kako bi se povećao nivo kvaliteta podataka u dnevniku. U jednoj primarnoj studiji, blokčejn tehnologija je primenjena kako bi se očistile netačne vremenske odrednice i nazivi aktivnosti pomoću pametnih ugovora (engl. *smart contracts*) [133].

Tabela 4. Grupisane tehnike čišćenja dnevnika izvršenja događaja prema literaturi

Kategorije	Tehnike	Softverski alat	Učestalost primene	Primarne studije
Klasterovanje putanja	Dodatak za klasterovanje putanja	ProM	13%	[114–119]
	Minimalno drvo raspona klasterovanje	ProM	3%	
	Klasterovanje zasnovano na statističkoj inferentnosti	/	3%	
Ukupno			19%	
Filtriranje putanje/događaja	Algoritam grananja i spajanja	/	9%	[116,120–127]
	Filtriranje aktivnosti zasnovano na entropiji	RapidProm	3%	
	Filter retkog ponašanja	ProM	6%	
	Dodatak za popravku dnevnika	ProM	6%	
Ukupno			24%	
Veštačka inteligencija, Mašinsko učenje, Duboko učenje	Bajezove mreže	Matlab	9%	[105–113]
	Algoritam nasumične šume	/	3%	
	Algoritmi SIER i MIEC	/	3%	
	Prirodna obrada jezika	/	6%	
	LSTM veštačke neuronske mreže	/	3%	
	CART algoritam	/	3%	
Ukupno			27%	
Tehnike popravke dnevnika izvršenja događaja	Dodatak za heurističko popravljjanje dnevnika	ProM	6%	[128,129]
Ukupno			6%	
Ugrađena obrada podataka	Induktivni algoritam	ProM	3%	[130]
	ILP algoritam	ProM	3%	
Ukupno			6%	
Tehnike poravnavanja	Poravnanje zasnovano na trošku	ProM	6%	[105,107,131]
	Provera usaglašenosti zasnovana na poravnanju	ProM	3%	
Ukupno			9%	
Apstrakcija događaja	Semantička apstrakcija	CPNTools	3%	[132]
Drugo	Blokčejn tehnologija	/	3%	[133]

2.3. Rezultati prethodnih istraživanja

Naredna potpoglavlja predstavljaju pregled prethodnih istraživanja na temu metodologija i uputstava za primenu automatskog otkrivanja poslovnih procesa u industriji, kao i prethodne preglede literature u oblasti.

2.3.1. Postojeće metodologije i uputstva

Kako bi se stekao uvid u prethodne metodologije i uputstva za primenu automatskog otkrivanja poslovnih procesa u industriji, izvršen je pregled literature. Analiza rezultata je pokazala da se prethodna istraživanja na ovu temu mogu grupisati na osnovu oblasti primene.

Prva grupa predstavlja metodologije iz oblasti automatskog otkrivanja poslovnih procesa koje su opšteg karaktera, odnosno primenljive su u bilo kojoj industriji i sa bilo kojim projektnim ciljem [15–20]. Druga grupa predstavlja specifične metodologije koje su fokusirane na primenu automatskog otkrivanja poslovnih procesa u zdravstvenoj industriji [21–25]. Treća grupa predstavlja uputstva koja imaju za cilj da olakšaju odabir algoritama za otkrivanje poslovnih procesa [134–138].

Jedna od prvih metodologija automatskog otkrivanja poslovnih procesa je imala za cilj da pruži širok pregled poslovnog procesa za kratak vremenski period, bez poznavanja domena u kojem se proces odvija [19]. Autori su naveli 6 faza projekata automatskog otkrivanja poslovnih procesa, bez posebnog fokusa na konkretne aktivnosti i algoritme/tehnike koje trebaju da se primene, osim u fazi analize uloga poslovnih resursa. Iste godine, predstavljena je metodologija za redizajn poslovnih procesa, koja je koristila automatsko otkrivanje poslovnih procesa i simulaciju radi predviđanja ponašanja procesa [20]. Metodologija je primenljiva na nestruktuiranim procesima i predstavlja korisno rešenje za kombinaciju simulacije i automatskog otkrivanja poslovnih procesa, iako su faze metodologije i aktivnosti definisane tek kroz studiju slučaja, i ne uključuju moguće algoritme i tehnike.

Najuticajniji autor iz oblasti automatskog otkrivanja poslovnih procesa Vil Van der Alst, sa Tehničkog Univerziteta u Ajndhovenu, definisao je L* metodologiju za unapređenje struktuiranih procesa zasnovanoj na praktičnoj primeni automatskog otkrivanja poslovnih procesa [17]. Doprinos ove metodologije je definisan životni ciklus projekata automatskog otkrivanja poslovnih procesa uz 5 ključnih faza. Metodologija obuhvata sve tipove i perspektive automatskog otkrivanja poslovnih procesa, uz objašnjenje pravilne implementacije. Mogući algoritmi i tehnike su takođe navedeni za svaku fazu, iako nije naglašeno kako odabrati algoritam ili tehniku u različitim situacijama. Iste godine, takođe na Tehničkom Univerzitetu u Ajndhovenu, objavljen je master rad kojim je predstavljena metodologija automatskog otkrivanja poslovnih procesa koja je nezavisna od industrije, softverskih alata i primene [18]. Metodologija je razvijena pomoću pristupa sistemskog inženjerstva i uključuje 4 faze. Doprinos metodologije je sveobuhvatan pregled svih koraka i aktivnosti koje je potrebno sprovesti, iako nisu navedeni alati, algoritmi i tehnike koji se mogu primeniti u različitim fazama projekta.

Metodologija koja ima za cilj unapređenje performansi poslovnih procesa i usaglašenost poslovnih procesa sa poslovnim pravilima i regulacijama se naziva projektna metodologija automatskog otkrivanja poslovnih procesa (engl. *Process Mining Project Methodology* – PM²) [16]. Metodologija je primenljiva za analizu struktuiranih i nestruktuiranih poslovnih procesa i sačinjena je od 6 faza. Prednost ove metodologije su definisani ulazi i izlazi za svaku fazu, koji pomažu praktikantima da razumeju potrebe i očekivane rezultate svake faze. Takođe, u fazi procesuiranja podataka, autori su detaljno predstavili različite tehnike i aktivnosti koje treba da se sprovedu. U fazi planiranja projekata automatskog otkrivanja poslovnih procesa, autori su naglasili da karakteristike samog poslovnog procesa i kvalitet dnevnika izvršenja događaja utiču na odabir algoritama i tehnika, primenjenih u daljem toku izvršenja projekta. Nije specificirano kako ovi elementi projekata utiču na proces selekcije.

Poslednja metodologija automatskog otkrivanja poslovnih procesa opšteg karaktera ima veoma praktičnu primenu. Metodologija je dizajnirana kroz formalnu inženjersku metodu tako što su anketirani vodeći stručnjaci iz oblasti automatskog otkrivanja poslovnih procesa o njihovim očekivanjima u pogledu metodologije automatskog otkrivanja poslovnih procesa [15]. Doprinos metodologije su tri

studije slučaja koje se razlikuju prema karakteristikama dnevnika izvršenja događaja i poslovnim procesima. Opisane studije slučaja uključuju informacije o ciljevima, izvorima podataka, algoritmima, tehnikama i alatima primenjenim u svakom koraku izvršavanja projekta.

Pored metodologija koje su opšteg karaktera, u literaturi se javlja i nekoliko metodologija koje su primenljive isključivo u zdravstvenoj industriji [21–25]. Neke od ovih metodologija objašnjavaju detaljno kako sprovesti aktivnosti projekta različitog nivoa detaljnosti događaja, kao i koje algoritme i tehnike primeniti na dnevnik izvršenja događaja u različitim situacijama. Međutim, ove metodologije i uputstva su primenljivi isključivo na poslovnim procesima i dnevnicima izvršenja događaja zdravstvene industrije, koja koristi specifične softverske alate i ima specifične ciljeve.

Treća grupa primarnih studija je sačinjena od uputstava koja imaju za cilj da olakšaju odabir algoritama za otkrivanje poslovnih procesa [134–138]. Sva uputstva stavljaju fokus na metapodatke dnevnika izvršenja događaja, na koje se primenjuju različiti algoritmi i statističke metode kako bi se utvrdilo koji algoritam za otkrivanje poslovnih procesa kreira poslovne modele najvišeg kvaliteta. Ovakav pristup zahteva da prilikom planiranja projekta AOPP-a praktikant treba da poznaje tehnike statističke analize i programiranja, što često nije slučaj. Jedan od ciljeva IEEE radne grupe za automatsko otkrivanje poslovnih procesa jeste da se poveća upotrebljivost i razumljivost automatskog otkrivanja poslovnih procesa i omogući primena od strane menadžera i analitičara poslovnih procesa.

U literaturi postoji i rad koji se karakteriše kao uputstvo za praktikante automatskog otkrivanja poslovnih procesa [139]. Autori su izvršili sistematski pregled literature preko 800 radova iz oblasti automatskog otkrivanja poslovnih procesa, kako bi identifikovali glavne slučajeve upotrebe i istraživačka pitanja koja se postavljaju unutar određenih slučajeva upotrebe. Autori nisu beležili podatke vezano za industriju, softverske alate i primenjene konkretne tehnike različitih tipova automatskog otkrivanja poslovnih procesa. Takođe, nije ispitivana veza između slučajeva upotrebe i definisanih pitanja u okviru njih.

Nakon analize postojećih metodologija i uputstava za izvođenje projekata automatskog otkrivanja poslovnih procesa i selekciju tehnika i algoritama, može se zaključiti da u literaturi postoje definisane faze projekata, kao i aktivnosti koje se sprovode u okviru projekata. Većina metodologija nije pružila praktična uputstva za implementaciju koraka, kao ni uputstva za odabir različitih tehnika i algoritama. Praktikanti automatskog otkrivanja poslovnih procesa mogu da koriste objavljene studije slučaja ili da koriste neko od postojećih rešenja za odabir jedne vrste algoritma, za koje je potrebno poznavanje statističkih metoda i programiranja. Postoji jasna potreba za istraživanjem koje će utvrditi faktore koji utiču na odabir tipova, perspektiva, algoritama, tehnika i softverskih alata za izvođenje automatskog otkrivanja poslovnih procesa, a koje ne zahteva višedomensko znanje praktikanata.

2.3.2. Prethodni pregledi literature

S obzirom da je automatsko otkrivanje poslovnih procesa mlada oblast sa metodologijama, smernicama i novim pristupima koji se još uvek razvijaju, postoje relevantni pregledi primene automatskog otkrivanja poslovnih procesa u različitim domenima, sa ciljem da se sintetizuje trenutno znanje. Tabela 5 predstavlja preglede istaknute literature iz oblasti automatskog otkrivanja poslovnih procesa od 2008. do 2023. godine, poređane hronološki.

Pregledi se mogu podeliti na deskriptivne preglede literature, sistematske preglede literature i sistematske studije mapiranja. Teme pregleda literature variraju od određenih problema, algoritama i tehnika, industrijskih aplikacija do pregleda u određenim domenima, kao što su simulacija, obrazovanje, razvoj softvera i zdravstvena zaštita. Rani pregledi literature imali su za cilj da uvedu automatsko otkrivanje poslovnih procesa kao disciplinu, fokusirajući se na koncepte, dostupne algoritme otkrivanja, probleme i trendove [140–144]. Od 2014. godine fokus se pomerio na primenu u specifičnim okruženjima i domenima, kao što su simulacija procesa [145] i nauka o podacima [146], agilni razvoj softvera [147], obrazovanje [148], lanci snabdevanja [149] i primena u IKT industriji [150,151]. Štaviše, od 2018. godine, većina pregleda literature analizirala je širi spektar tehnika [152–154], poboljšane pristupe automatskom otkrivanju poslovnih procesa [155,156] i upravljanje ljudskim resursima [157]. Primena automatskog otkrivanja poslovnih procesa je tema u skoro trećini svih pregleda literature [158–

164]. Prema pregledima literature, prednosti u zdravstvu su otkrivanje modela procesa, provera usaglašenosti kliničke stvarnosti sa medicinskim smernicama i pronalaženje mogućnosti za poboljšanje i definisanje najboljih praksi različitih medicinskih procesa. Takođe, pregledi su obuhvatili podatke o analiziranim medicinskim procesima i tipovima podataka, istraživačkim pitanjima ili ciljevima i primenjenim tehnikama, perspektivama i softverskim alatima.

Ostali pregledi literature su se bavili sa primenom u industriji uopšte, sa posebnim fokusom na specifične poslovne procese, tehnike i algoritme [27,139,165–167]. Takođe, jedan sistematski pregled literature je izvršio sintezu tehnika čišćenja dnevnika izvršenja događaja [14].

Nijedan od predstavljenih pregleda literature nije statistički obrađivao podatke o primeni u industriji kako bi se ispitale veze između ključnih faktora koji određuju tok izvršavanja analize procesa primenom automatskog otkrivanja poslovnih procesa.

Tabela 5. Prethodni pregledi literature iz oblasti automatskog otkrivanja poslovnih procesa

Pristup	Tema	Godina	Primarna studija
Deskriptivni pregled literature	Trendovi i problemi	2008	[140]
Deskriptivni pregled literature	Algoritmi za otkrivanje poslovnih procesa	2009	[141]
Deskriptivni pregled literature	Algoritmi za otkrivanje poslovnih procesa	2011	[142]
Deskriptivni pregled literature	Osnovni koncepti i primena	2012	[143]
Deskriptivni pregled literature i komparativna analiza	Algoritmi za otkrivanje poslovnih procesa	2014	[144]
Deskriptivni pregled literature	Simulacija	2014	[145]
SPL	Primena u zdravstvu	2016	[158]
SPL	Primena u zdravstvu	2016	[159]
SPL	Primena u zdravstvu	2016	[160]
SPL	Primena u obrazovanju	2017	[148]
SPL	Agilni razvoj softvera	2018	[147]
SPL	Primena u industriji	2018	[4]
SPL	Tehnike unapređenja poslovnih procesa	2018	[154]
SPL	Primena u zdravstvu	2018	[161]
Sistematsko mapiranje	Tehnike	2018	[152]
SPL	Primena u industriji	2018	[167]
Sistematsko mapiranje	Upravljanje ljudskim resursima	2018	[157]
SPL	Primena u lancima snabdevanja	2018	[149]
SPL	Primena u zdravstvu	2018	[162]
SPL	Primena u zdravstvu	2019	[163]
SPL	Primena u računarstvu	2019	[150]
Sistematsko mapiranje	Tehnike	2019	[153]
Deskriptivni pregled literature	Primena u industriji	2019	[166]
SPL	Nauka o podacima	2019	[146]
SPL	Otkrivanje poslovnih procesa vođeno ciljevima	2020	[155]
SPL	Dupli zadaci	2020	[156]
SPL	Primena u industriji	2020	[165]
Deskriptivni pregled literature	Primena u zdravstvu	2020	[164]
SPL	Primena u proizvodnoj industriji	2020	[27]

SPL	Tehnike čišćenja dnevnika izvršenja događaja	2021	[14]
SPL	Primena u sigurnosti na internetu	2022	[151]
Sistematsko mapiranje	Primena u zdravstvu	2022	[168]
SPL	Slučajevi upotrebe	2022	[139]

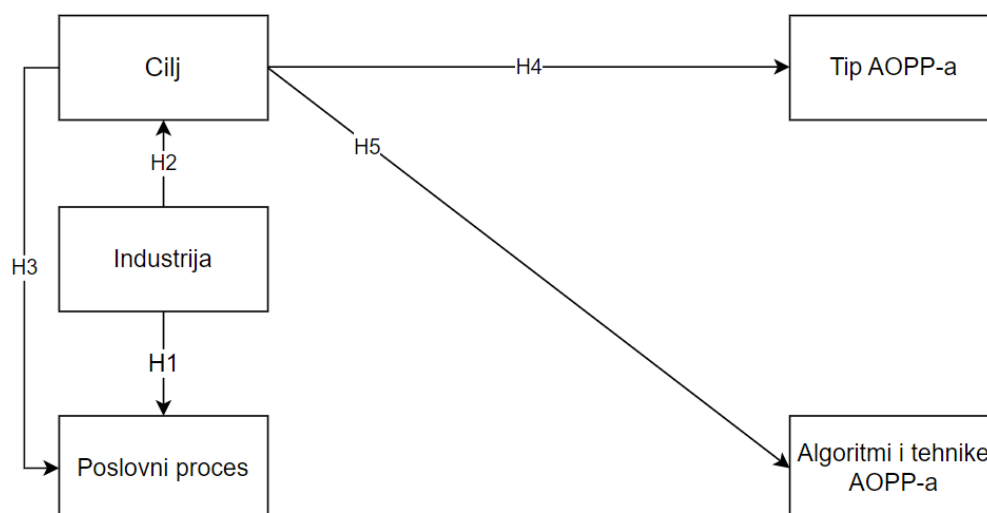
2.4. Istraživački modeli i hipoteze

Na osnovu predstavljenih pregleda literature koja se tiču dva dela istraživanja doktorske disertacije, autor definiše dva zasebna istraživačka modela i odgovarajuće hipoteze, kako bi se odgovorilo da postavljena istraživačka pitanja.

Kao što je definisano u potpoglavlju 1.2., prvo istraživačko pitanje glasi:

IP1: Koji faktori projekata analize i automatskog otkrivanja poslovnih procesa (projektne karakteristike) utiču na odabir tipova, softverskih alata i algoritama/tehnika automatskog otkrivanja poslovnih procesa?

Slika 10 predstavlja odgovarajući istraživački model 1 (IM1) sa hipotezama o vezama između ključnih faktora projekata automatskog otkrivanja poslovnih procesa.



Slika 10. Istraživački model 1

Ključni činioci istraživanja koji daju odgovor na IP1 će činiti istraživački model 1 i to su industrija, tip poslovnog procesa, cilj analize, tipovi automatskog otkrivanja poslovnih procesa i algoritmi i tehnike, izuzev tehnika čišćenja dnevnika izvršenja događaja.

U narednom tekstu opisani su ključni činioci istraživanja unutar IM1 i teorijske osnove za definisanje njihovih veza.

Povezanost faktora industrija i poslovni proces

Prethodno sprovedeni sistematski pregledi literature predstavljeni u potpoglavljima 2.2.1 i 2.2.2. pokazali su da različite industrije primenjuju različite tipove i perspektive AOPP-a [4,27,28]. Na primer, zdravstvena industrija se razlikuje od ostalih po procentu studija slučaja koje su analizirale proveru usaglašenosti poslovnog procesa. Detaljnijom analizom je utvrđeno da svaka industrija (npr. zdravstvena, proizvodna, obrazovna) analizira poslovne procese koji su specifični za tu industriju. Takođe, poslovni procesi se razlikuju prema broju aktivnosti koje obuhvataju, prema učestalosti pojave tih aktivnosti, prema broju varijanti izvršenja procesa, kao i prema nivou složenosti modela koji ih predstavlja. Različite vrste algoritama bi trebale da se primene na različite vrste poslovnih procesa radi

otkrivanja modela procesa [4,27,28]. Postavlja se pretpostavka da postoji statistički značajna povezanost između vrste poslovnog procesa i industrije.

Povezanost faktora industrija i cilj analize

S obzirom da analiziraju različite vrste poslovnih procesa, industrije mogu imati i različite ciljeve analize [146,148,151,160,165]. S tim u vezi, moguće je pretpostaviti da postoji statistički značajna povezanost između industrija i ciljeva, prilikom primene automatskog otkrivanja poslovnih procesa.

Povezanost faktora cilj i poslovni proces

S obzirom na specifičnost konteksta u kojem se neki poslovni proces odvija, može se pretpostaviti da različite vrste poslovnih procesa mogu imati različite ciljeve analize prilikom izvršavanja automatskog otkrivanja poslovnih procesa [30,72,169–171]. Takođe, na osnovu prethodnih istraživanja je utvrđeno da postoji potencijalna veza između cilja analize automatskog otkrivanja poslovnih procesa i vrste poslovnog procesa [4,16,172,173]. Postavlja se pretpostavka da postoji statistički značajna povezanost između ciljeva analize i vrste poslovnog procesa koji se analizira.

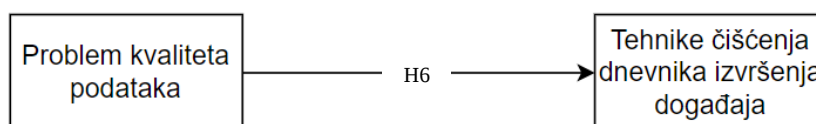
Povezanost faktora cilj analize sa faktorima tip, algoritmi i tehnike AOPP-a

Ne postoji prethodno istraživanje o uticaju ciljeva analize na odabir koraka i tehnika AOPP-a. Međutim, u oblasti koja je bliska AOPP-a, istraživanje i eksploatacija podataka, autori su zaključili da na odabir tehnika i algoritama najviše utiču ciljevi i metapodaci analiziranog seta podataka [172,173]. Metodologije AOPP-a preporučuju postavljanje eksplicitnih ciljeva kroz istraživačka pitanja, ključne indikatore performansi ili definisanje problema koje AOPP-a treba da reši [16,17,155]. Postavlja se pretpostavka da postoji statistički značajna veza između cilja analize sa primenjenim tipom automatskog otkrivanja poslovnih procesa, kao i algoritmima i tehnikama.

Dalje, kao što je definisano u potpoglavlju 1.2., drugo istraživačko pitanje glasi:

Koje tehnike čišćenja treba da se primene na dnevnik izvršenja događaja u zavisnosti od problema kvaliteta podataka?

Slika 11 predstavlja odgovarajući istraživački model 2 (IM2) sa hipotezom.



Slika 11. Istraživački model 2

Povezanost problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja

U oblastima bliskim automatskom otkrivanju poslovnih procesa postoji pretpostavka, zasnovana na iskustvu i domenskom znanju praktikanata, da se u zavisnosti od vrste problema kvaliteta mogu primeniti različite grupe tehnika za čišćenje podataka i uklanjanje problema [174,175]. Međutim, ta pretpostavka nije formalno ispitana. U okviru istraživačkog modela 2 biće ispitana povezanost problema kvaliteta i tehnika čišćenja dnevnika, čije su dimenzije definisane na osnovu izvršenih pregleda literature i sinteze problema kvaliteta i tehnika izvršenja događaja. Definisane ključne stavke koje određuju činioce istraživanja u IM2 predstavljene su u tabeli 6.

Tabela 6. Stavke koje određuju dimenzije ključnih činilaca istraživanja u okviru IM2

Dimenzija	Stavka	Izvor
Problemi kvaliteta podataka	Nedostajući podaci: Slučaj	[7]
	Nedostajući podaci: Događaj (Raštrkani događaj)	[7,8]
	Nedostajući podaci: Veza (Neuhvatljiv slučaj)	[7,8]
	Nedostajući podaci: Naziv aktivnosti	[7]
	Nedostajući podaci: Atribut slučaja/događaja	[7]
	Nedostajući podaci: Vremenska odrednica	[7]
	Nedostajući podaci: Resurs	[7]
	Netačni podaci: Slučaj	[7]
	Netačni podaci: Događaj	[7]
	Netačni podaci: Veza (Raštrkani slučaj)	[7,8]
	Netačni podaci: Naziv aktivnosti (Zagađeni/Iskrivljeni naziv atributa)	[7,8]
	Netačni podaci: Atribut slučaja/događaja	[7]
	Netačni podaci: Vremenska odrednica (Hvatanje događaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj)	[7,8]
	Netačni podaci: Resurs (Zagađeni naziv atributa)	[7,8]
	Neprecizni podaci: Veza	[7]
	Neprecizni podaci: Naziv aktivnosti (Istoimeni naziv atributa)	[7,8]
	Neprecizni podaci: Atribut slučaja/događaja (Sinonimni naziv atributa)	[7,8]
	Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj)	[7,8]
	Neprecizni podaci: Resurs	[7]
	Irelevantni podaci: Slučaj	[7]
	Irelevantni podaci: Događaj (Hvatanje događaja zasnovano na obrascu, Kolateralni događaj)	[7,8]
	Količina podataka, nivo detaljnosti događaja, kompleksnost	[7]
Tehnike čišćenja dnevnika izvršenja događaja	Klasterovanje putanja	[14,114–119]
	Tehnike popravke dnevnika izvršenja događaja	[14,128,129]
	Filtriranje putanja/događaja	[14,116,120–127]
	Apstrakcija događaja	[14,132]
	Veštačka inteligencija, mašinsko učenje, duboko učenje	[14,105–113]
	Tehnike zasnovane na poravnanju	[14,105,107,131]
	Ugrađena obrada podataka	[130]

3. METODOLOŠKI ASPEKTI ISTRAŽIVANJA

Nakon definisanja dizajna istraživanja kroz definisane istraživačke probleme, utvrđivanje potrebe za istraživanjem i pregled ključnih čionioaca istraživanja, potrebno je pažnju preusmeriti na metodološke aspekte istraživanja. Metodološki aspekti istraživanja podrazumevaju opis metodologije sprovođenja istraživanja, definisanje instrumenta za prikupljanje podataka i predstavljanje plana obrade tih podataka, koji su uslovljeni samim dizajnom istraživanja [176,177]. Kako bi dalje razumeo opis razvoja mernih instrumenata i prikupljanja podatka, potrebno je definisati skale merenja koje određuju tip varijable. Merne skale se mogu grupisati prema sledećim nivoima merenja [177–179]:

1. Nominalni nivo je nivo merenja u okviru kojeg se vrši svrstavanje članova uzorka prema određenoj karakteristici, u dve ili više kategorija. Objektima se pridružuje broj kada oni poseduju istu osobinu. Osobine se mogu međusobno porediti, ali se ne zna njihova tačna razlika. Varijable koje se mere nominalnom skalom se nazivaju kategoričke varijable.

2. Ordinalni nivo merenja se koristi u situaciji kada niz članova uzorka raste ili opada prema zastupljenosti posmatrane osobine. Objektima se pridružuju brojevi tako da poredak brojeva održava poredak definisan relacijom poređenja prisutnosti osobine.

3. Intervalni nivo merenja podrazumeva svrstavanje članova uzorka u opadajući ili rastući niz po određenom mernom svojstvu, tako da su rastojanja između članova poznata. Objektima se pridružuju brojevi na način da razlike između tih brojeva reflektuju razlike u stepenu zastupljenosti posmatrane osobine. Intervalni nivo poseduje proizvoljnu nultu tačku i jedinicu merenja veličine svojstva.

4. Nivo razmere predstavlja skalu po osnovu koje se objektima pridružuju brojevi tako da razlike između njih reflektuju jasnu razliku između vrednosti osobine. Tačnu razliku je moguće izmeriti jer racio skala poseduje apsolutnu nulu i jedinicu merenja na osnovu koje se može izmeriti količina svojstva, tj. osobine.

S obzirom da doktorska disertacija sadrži dva konkurentna istraživačka modela, koji zahtevaju dve različite metodologije prikupljanja i obrade podataka, ovo poglavlje se dalje deli na dve odgovarajuće celine.

3.1. Metodologija prikupljanja i obrade podataka istraživačkog modela I

Istraživački model 1 (IM1) se odnosi na ispitivanje veze između ključnih faktora projekata automatskog otkrivanja poslovnih procesa, koji su sprovedeni nad stvarnim poslovnim procesima u različitim industrijama. Stoga je način prikupljanja podataka za ovaj istraživački model sistematski pregled literature objavljenih studija slučaja sa temom primene automatskog otkrivanja poslovnih procesa u stvarnim poslovnim sistemima.

Sistematski pregled literature je sačinjen od tri faze: planiranje, izvođenje i izveštavanje.

Planiranje pregleda literature ima za cilj da definiše protokol pregleda literature. Protokol pregleda literature sadrži definisanu potrebu za pregledom literature, istraživačka pitanja, ciljne baze podataka, termine pretrage, kriterijume inkluzije i ekskluzije, kao i strategiju prikupljanja podataka [180]. Istraživačka pitanja treba da budu jasno definisana, tako da dovedu do značajnih rezultata koji imaju uticaj na oblast istraživanja [180]. Nakon definisanja istraživačkih pitanja pregleda literature, utvrđuje se strategija pretrage primarnih studija. Prvi korak definisanja strategije pretrage podrazumeva inicijalni pregled radova na temu, kako bi se što tačnije utvrdili termini za pretragu i indeksne baze podataka koje će biti pretraživane. Naredni korak je postavljanje kriterijuma pretrage, koji mogu biti formulisani kao kriterijumi uključivanja i isključivanja radova iz sistematskog pregleda literature. Takođe, poželjno je postaviti kriterijume kvaliteta koje primarne studije moraju da ispune kako bi bile uključene u istraživanje. Kriterijumi kvaliteta nisu jasno određeni u literaturi, ali postoji uopštena preporuka da se kvalitet odnosi na meru primenljivosti rezultata studije, kao i meru u kojoj dizajn i sprovođenje studije može da dovede do sistemske greške u rezultatima [180]. Poslednji korak definisanja

protokola pregleda literature jeste utvrđivanja strategije prikupljanja podataka, gde je potrebno utvrditi varijable koje će se posmatrati i njihove dozvoljene vrednosti.

Nakon što je definisan protokol istraživanja, započinje se sa izvođenjem pregleda literature. Aktivnosti koje se sprovode u okviru izvođenja pregleda literature su: identifikacija primarnih studija, selekcija primarnih studija, provera kvaliteta, ekstrakcija podataka i sinteza podataka [180]. Faza izveštavanja podrazumeva objavljivanje dobijenih rezultata kroz tehnički izveštaj, doktorsku disertaciju, kao rad u časopisu ili na konferenciji [180].

3.1.1. Protokol pregleda literature

Potreba za sistematskim pregledom literature kao sredstvom za prikupljanje podataka je utvrđena prilikom opisa motivacije istraživanja u potpoglavlju 1.1 i prilikom uvida u prethodne preglede literature u potpoglavlju 2.3.2. Istraživačka pitanja sistematskog pregleda literature, nastala na osnovu IP1, glase:

- Koji su ključni činioci projekata automatskog otkrivanja poslovnih procesa?
- Da li postoji veza između ključnih činioca projekata automatskog otkrivanja poslovnih procesa?

Indeksne baze podataka koja su pretraživane su *Scopus*, *Web of Science*, *EBSCO* i *IEEE Explore*. Termin pretrage koji je korišćen za pretragu među naslovima, apstraktima i ključnim rečima je sledeći:

(„process mining“) AND („case study“ OR application) AND PUBYEAR > 2018 AND
PUBYEAR < 2024

Radovi su morali biti napisani na engleskom jeziku i objavljeni kao članak u časopisu, rad na konferenciji ili poglavlje knjige. Prilikom inicijalnog pregleda radova, izvršeno je čitanje apstrakata, gde su postavljani razlozi na osnovu kojih rad može biti odbačen. Razlozi su sledeći:

- Razlog 1: Rad mora da predstavlja primenu automatskog otkrivanja poslovnih procesa u industriji. Ukoliko rad predstavlja novu metodologiju ili alat, njihova primena mora biti prikazana kroz studiju slučaja.
- Razlog 2: Rad ne sme da prikazuje samo primenu tehnika čišćenja dnevnika izvršenja događaja, bez prikazanih ostalih koraka projekata automatskog otkrivanja poslovnih procesa.
- Razlog 3: Rad ne sme da prikazuje isključivo primenu sličnih tehnologija, poput mašinskog učenja, bez da su primenjene i tehnike automatskog otkrivanja poslovnih procesa.
- Razlog 4: Ukoliko neka grupa autora objavi više radova sa prikazanom istim rezultatima istraživanja (studijom slučaja), prihvata se samo jedan, aktuelniji rad.

Zatim, kriterijumi prihvatanja radova koji se primenjuju prilikom čitanja celih radova su sledeći:

- Kriterijum 1: Rad mora da sadrži podatke o vrsti analiziranog poslovnog procesa.
- Kriterijum 2: Rad mora da sadrži podatke o cilju analize.
- Kriterijum 3: Rad mora da izveštava o primenjenim tipovima i tehnikama automatskog otkrivanja poslovnih procesa.
- Kriterijum 4: Primena automatskog otkrivanja poslovnih procesa mora da bude sprovedena u stvarnom poslovnom okruženju, bez analize veštačkih setova podataka.

Kriterijum kvaliteta određen u skladu sa literaturom [180,181] podrazumeva da su prihvaćene primarne studije izvedene u skladu sa jasno definisanim ciljem i metodologijom, kao i da su predstavljeni rezultati jasni, sveobuhvatni i pouzdani.

Strategija prikupljanja podataka je definisana na osnovu ključnih činioca istraživačkog modela IM1, zajedno sa osnovnim bibliografskim podacima primarnih studija. Podaci su beleženi u MS Excel tabelu. Tabela 7 predstavlja ključne činioce istraživanja (varijable) koji su posmatrani kroz sistematski pregled literature, njihov tip podataka, mernu skalu i moguće vrednosti.

Tabela 7. Strategija prikupljanja podataka

Naziv varijable	Tip varijable	Merna skala	Moguće vrednosti
Naziv autora	Kategorička	Nominalna	/
Izvor	Kategorička	Nominalna	1 – rad na konferenciji 2 – rad u časopisu
Godina	Kategorička	Ordinalna	2019 - 2023
Industrija	Kategorička	Nominalna	1 – Proizvodnja 2 – Zdravstvo 3 – Obrazovanje 4 – Računovodstvo, kontroling, bankarstvo i osiguranje 5 – Informaciono-komunikacione tehnologije 6 – Logistika 7 – Javni servisi i usluge 8 – Građevina 9 – Energija
			1 – Otkrivanje i unapređenje poslovnog procesa 2 – Predviđanje ponašanja procesa 3 – Pronalazak anomalija i nepredviđenih aktivnosti 4 – Automatizacija poslovnog procesa 5 – Analiza resursa 6 – Analiza vremena čekanja 7 – Poređenje poslovnih procesa 8 – Provera usaglašenosti 9 – Otkrivanje poslovnih pravila 10 – Otkrivanje uzroka problema 11 – Analiza promena procesa
Cilj	Kategorička	Nominalna	1 – Montaža 2 – Upravljanje nepredviđenim okolnostima 3 – Nabavka 4 – Testiranje kvaliteta proizvoda

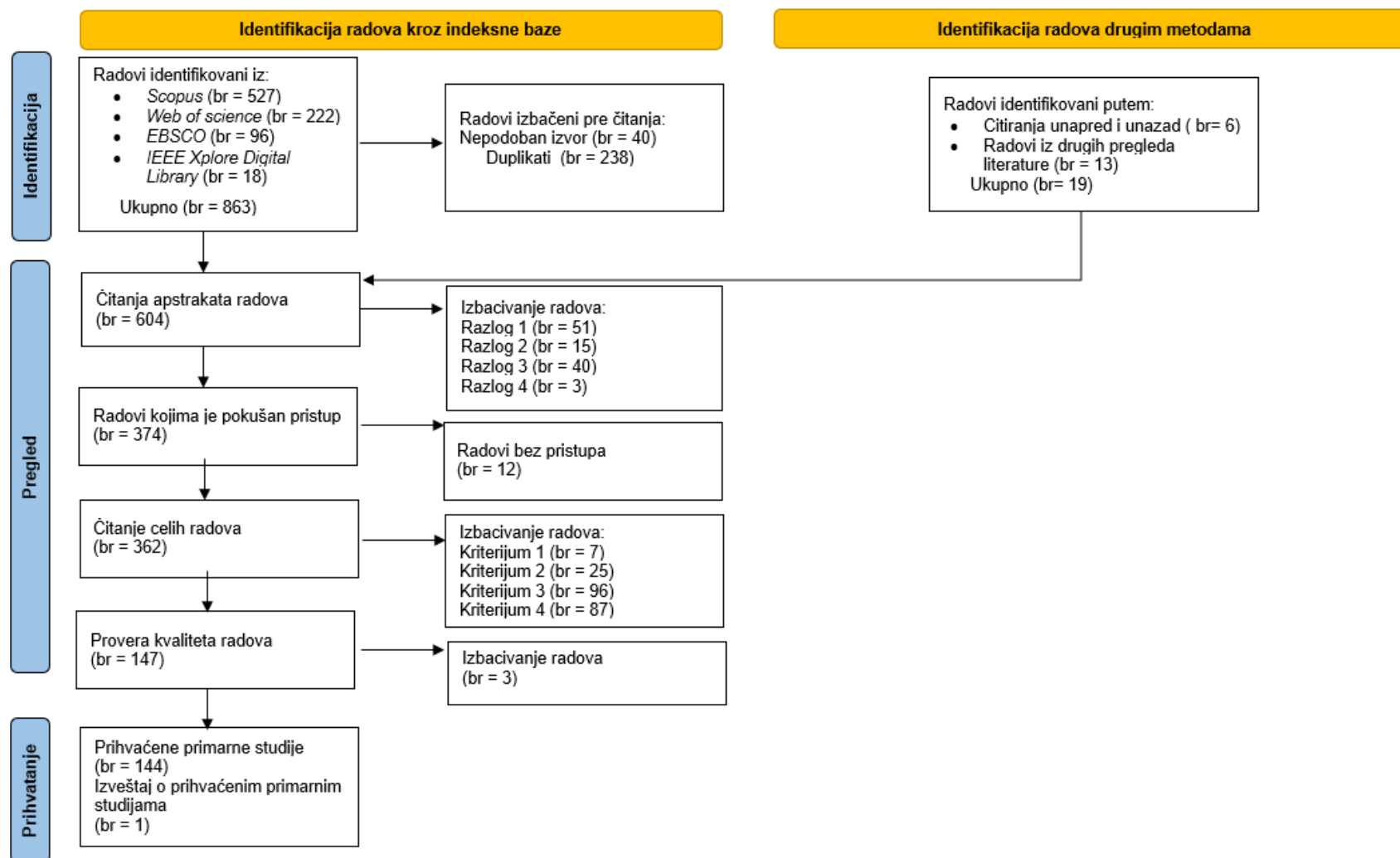
				5 – Planiranje proizvodnje itd.
Tip AOPP-a Otkrivanje poslovnog procesa Unapređenje poslovnog procesa Provera usaglašenosti Otkrivanje društvenih mreža Predviđanje	Kategorička	Nominalna	0 – nije primenjeno 1 – primenjeno	
Algoritam otkrivanja poslovnog procesa	Kategorička	Nominalna	1 – Fazi algoritam 2 – Heuristički algoritam 3 – Induktivni algoritam itd.	
Tehnike unapređenja	Kategorička	Nominalna	1 – Fazi algoritam ekstenzija 2 – Heuristički algoritam ekstenzija itd.	
Tehnike provere usaglašenosti	Kategorička	Nominalna	1 – Dodataka za proveru usaglašenosti 2 – Celonis provera usaglašenosti, itd.	
Tehnike predviđanja	Kategorička	Nominalna	1 – Algoritam mašinskog učenja 2 – Simulacija, itd.	
Tehnike analize društvenih mreža	Kategorička	Nominalna	1 – Dodatak za analizu društvenih mreža itd.	
Softverski alati	Kategorička	Nominalna	1 – <i>ProM</i> 2 – <i>Disco</i> 3 – <i>Celonis</i> , itd.	

3.1.2. Izvođenje pregleda literature

Izvođenja pregleda literature izvršeno je u skladu sa definisanim protokolom pregleda literature. Izveštavanje o toku izvođenja pregleda literature je u skladu sa *PRISMA* uputstvom za sprovođenje sistematskih pregleda literature [181], predstavljenim na slici 12.

Pokretanje termina pretrage kroz unapred utvrđene indeksne baze rezultiralo je sa pronalaskom 863 rada. Među tim radovima, na osnovu inicijalnog pregleda radova pre čitanja, pronađeno je 238 duplikata, kao i 40 radova koji su bili objavljeni kroz izvore koji nisu obuhvaćeni protokolom pregleda literature (radovi sa radionica, izveštaji, teze). Takođe, sprovedena je i manuelna pretraga radova koji su citirani od strane inicijalnog skupa radova, koji su citirali određene relevantne radove ili su uključeni u druge preglede literature na istu temu, koja je rezultirala sa dodatnih 19 radova. Naredni korak je bio čitanje apstrakata radova i provera definisanih razloga za odbacivanje radova. Glavni razlozi za odbacivanje radova u ovoj fazi su bili to što velik broj radova nije sadržao primenu novih alata ili pristupa u samoj industriji, kao i to što dosta istraživača uključuje termin „*process mining*” u ključne reči radova

koji primenjuje neke druge tehnike nauke o podacima na podatke o poslovnim procesima. Zatim je 362 rada pribavljeno radi čitanja celog teksta i primene kriterijuma prihvatanja radova. Radovi su najčešće odbacivani jer su predstavljali primenu automatskog otkrivanja poslovnih procesa nad sintetičkim podacima ili jer nisu dovoljno precizno izveštavali o samom postupku primene, pa nije bilo moguće zabeležiti relevantne podatke. Nakon odbacivanja radova koji nisu ispunjavali uslove, 147 radova je podleglo proveru kvaliteta, gde su odbačena dodatna 3 rada. Sprovođenje selekcije radova je rezultiralo sa 144 prihvaćene studije slučaja, iz kojih su izvučeni podaci za dalju obradu.



Slika 12. PRISMA tok izvršavanja identifikacije primarnih studija

3.1.3. Bibliografski podaci primarnih studija

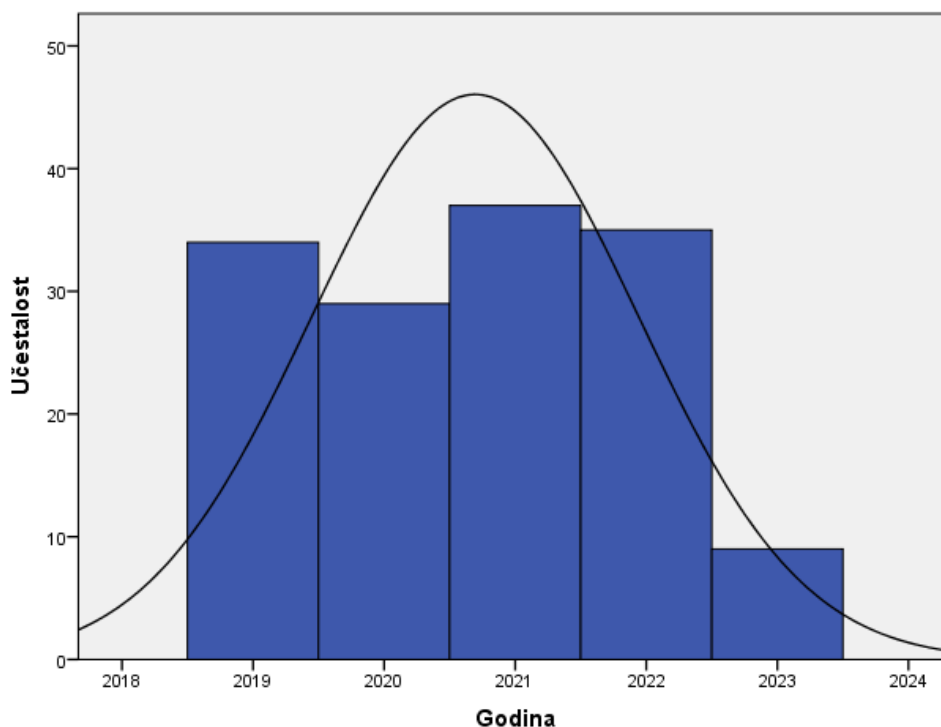
Kao što je prikazano na slici 12, primenom procedure sistematskog pregleda literature, prihvaćeno je 144 primarnih studija za dalju analizu. Spisak primarnih studija, podeljenih prema industriji u okviru koje je istraživanje sprovedeno i prema cilju sprovedene analize automatskog otkrivanja poslovnih procesa, priložen je kao Prilog A.

Primarne studije su objavljene u periodu od 2019. do 2023. godine. Tabela 8 predstavlja učestalost i procenat objavljenih primarnih studija po godinama.

Tabela 8. Učestalost i procenat primarnih studija po godinama

Godina	Učestalost	%
2019	34	23.6
2020	29	20.1
2021	37	25.7
2022	35	24.3
2023	9	6.3
Ukupno	144	100.0

Najveći broj primarnih studija je objavljen 2021. godine, sa udelom od 25.7% od svih primarnih studija. Zatim slede 2022. godina, sa udelom od 24.3% od svih primarnih studija, 2019. godina sa udelom od 23.6% od svih primarnih studija i 2020. godina sa udelom od 20.1% od svih primarnih studija. S obzirom na protokol sistematskog pregleda literature, gde je utvrđeno da je korak pretrage radova po utvrđenim terminima pretrage izvršen u martu 2023., najmanji broj primarnih studija je objavljen 2023. godine, sa udelom od 6.3%. Slika 13 predstavlja histogram objavljivanja primarnih studija po godinama sa krivom normalne raspodele.



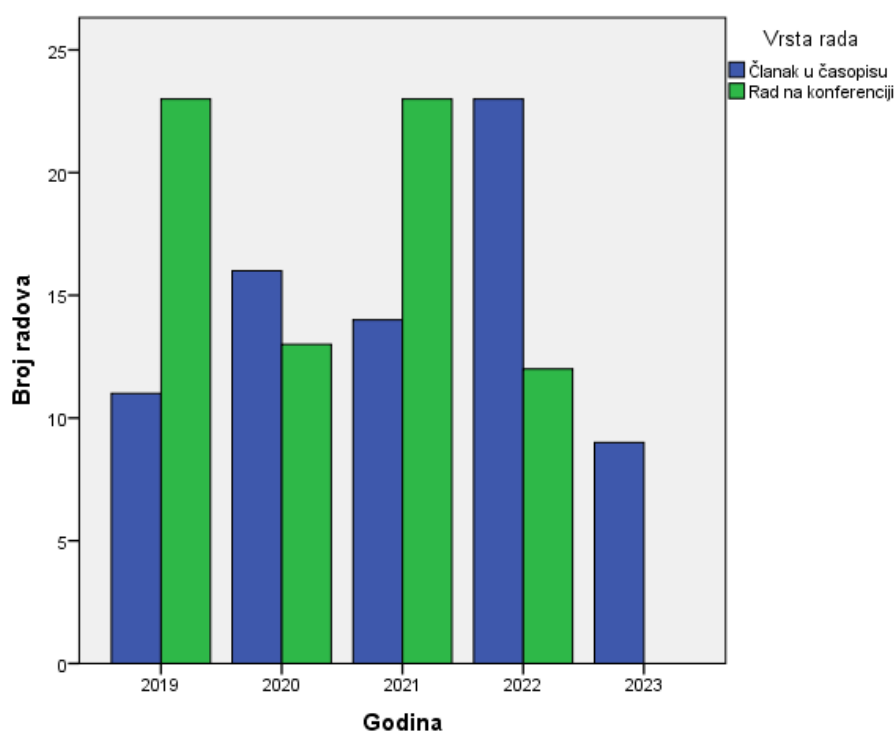
Slika 13. Histogram učestalosti objavljivanja primarnih studija po godinama

Primarne studije su objavljene kao radovi na konferenciji ili članak u časopisu. Tabela 9 predstavlja uvid u učestalost i procenat objavljenih primarnih studija prema vrsti rada. Zastupljenost članaka u časopisu i radova na konferenciji je skoro jednaka, sa 50.7% radova objavljenih kao članak u časopisu i 48.3% radova objavljenih kao rad na konferenciji.

Tabela 9. Učestalost i procenat primarnih studija prema vrsti rada

Vrsta rada	Učestalost	%
Članak u časopisu	73	50.7
Rad na konferenciji	71	49.3
Ukupno	144	100.0

Slika 14 predstavlja raspodelu radova u odnosu na vrstu rada i godinu objavljivanja. Može se zaključiti da postoji nekoliko trendova u pogledu njihovog odnosa, gde su 2019. i 2021. godine u značajno većoj meri istraživanja objavljivana na konferencijama, dok su u 2022. i 2023. godini istraživanja objavljivana u većoj meri kao članci u časopisima.



Slika 14. Broj objavljenih vrsta radova prema godinama

Članci u časopisima su objavljeni u 61 različitom časopisu. Tabela 10 predstavlja časopise u kojima je objavljeno više od jedne primarne studije, zajedno sa njihovom učestalošću i procentom u odnosu na ukupan broj primarnih studija u časopisima.

Tabela 10. Najučestaliji časopisi u kojima su objavljene primarne studije

Naziv časopisa	IF 2022	Učestalost	%
<i>International Journal of Health Planning and Management</i> (ISSN: 0749-6753)	2.7	5	6.8
<i>IEEE Access</i> (ISSN: 2169-3536)	3.9	3	4.1

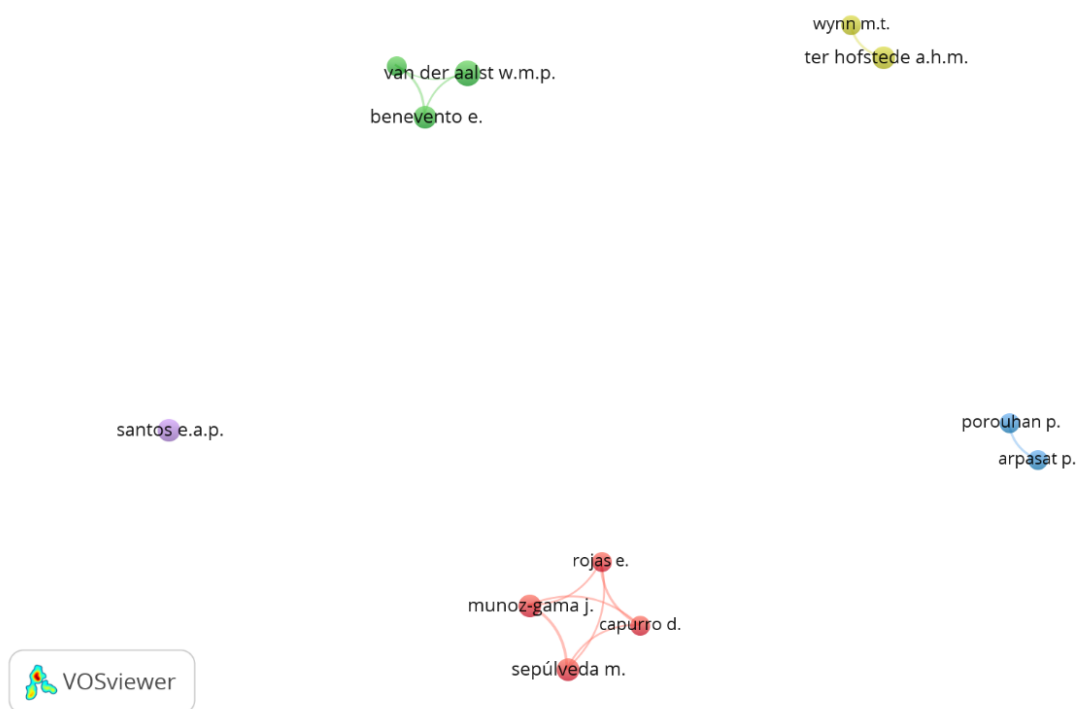
<i>Algorithms</i> (ISSN: 1999-4893)	2.3	3	4.1
<i>Expert Systems with Applications</i> (ISSN: 0957-4174)	8.5	3	4.1
<i>Applied Sciences</i> (ISSN: 2076-3417)	2.7	2	2.7
<i>BMC Medical Information and Decision Making</i> (ISSN: 1472-6947)	3.9	2	2.7
Drugi	/	55	75.3
Ukupno	/	73	100.0

Primarne studije su objavljene na 42 različite konferencije, sa 9 konferencija na kojima je objavljeno više od jednog rada. Tabela 11 predstavlja izvode sa konferencija na kojima je objavljeno više od jedne primarne studije.

Tabela 11. Naučestaliji izvodi konferencija u kojima su objavljene primarne studije

Naziv izvoda	Učestalost	%
<i>Lecture Notes in Business Information Processing</i>	12	16.9
<i>Lecture Notes in Computer Science</i>	5	7.0
<i>Proceedings of the International Conference on ICT and Knowledge Engineering</i>	5	7.0
<i>CEUR Workshop Proceedings</i>	4	5.6
<i>Advances in Intelligent Systems and Computing</i>	3	4.2
<i>ACM International Conference Proceedings Series</i>	3	4.2
<i>Journal of Physics: Conference Series</i>	2	2.8
<i>Proceedings of the Annual Hawaii International Conference on System Sciences</i>	2	2.8
<i>Proceedings of the LACCEI International Multiconference for Engineering, Education, and Technology</i>	2	2.8
Drugi	33	46.4
Ukupno	71	100.0

Kako bi se uvidelo ko su naznačajni autori koji istražuju primenu automatskog otkrivanja poslovnih procesa, izvršena je analiza saradnje autora u *VOSviewer* alatu za mapiranje bibliografskih podataka radova [182] i predstavljena na slici 15.



Slika 15. Mreža saradnje između autora primarnih studija

Radi kreiranja mreže saradnje autora, odabrani su istraživači koji su autori minimalno 3 primarne studije, gde je 12 autora odgovaralo kriterijumu. Autori su predstavljeni u tabeli 12. Autori su svrstani u 5 klastera, u zavisnosti od toga da li imaju zajedničke radove ili ne. Veličina čvora odgovara broju objavljenih radova, a debljina veze između autora odgovara učestalosti njihove saradnje.

Tabela 12. Grupisani autori primarnih studija sa najviše radova

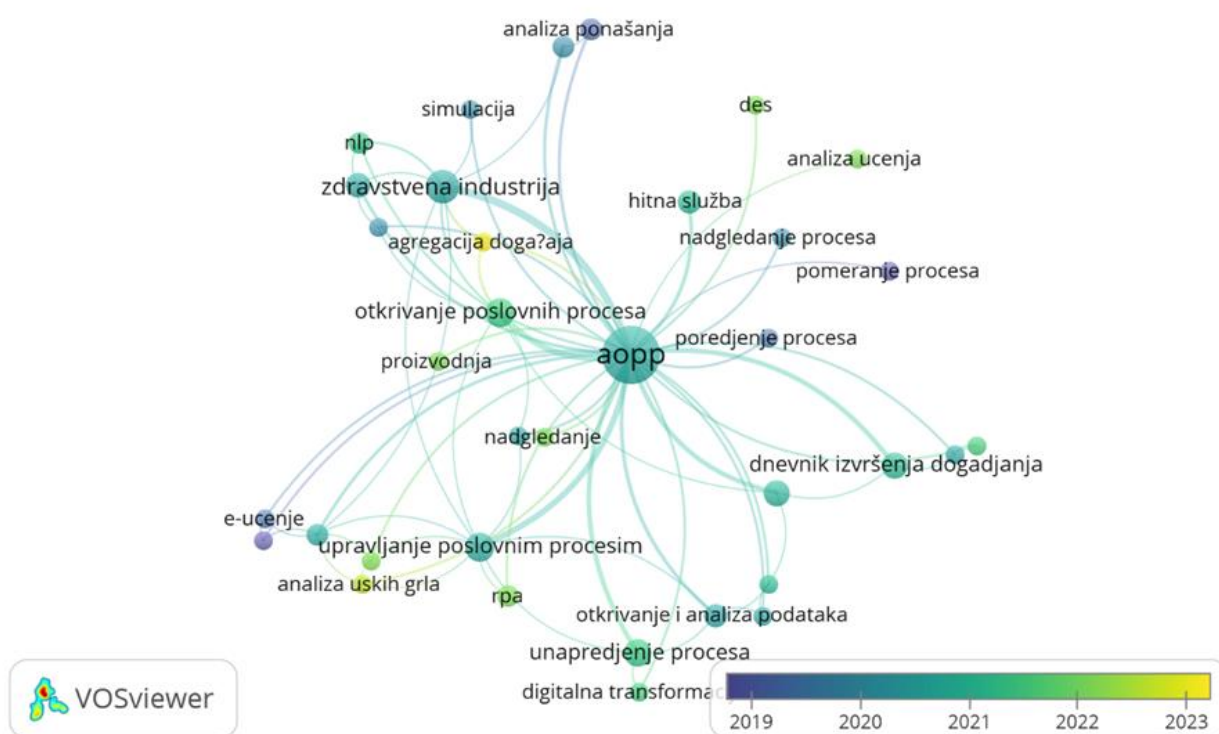
Autor	Broj radova	Klaster
Capurro, D.	3	1
Munoz-Gama, J.	4	1
Rojas, E.	3	1
Sepulveda, M.	4	1
Aloini, D.	3	2
Benevento, E.	4	2
Van der Aalst, W.	5	2
Arpasat, P.	3	3
Porouhan, P.	3	3
Ter Hofstede, A.H.M.	4	4
Wynn, M.T.	3	4
Santos, E.A.P.	4	5

Kako bi se analizirale prominentne teme u oblasti primene automatskog otkrivanja poslovnih procesa, kreirana je mreža povezanosti ključnih reči primarnih studija (engl. *keywords co-occurrence network*). Mreža povezanosti ključnih reči primarnih studija preuzetih zajedno sa radovima iz indeksnih baza *Scopus* i *Web of Science* kreirana je pomoću *VOSviewer* alata za mapiranje bibliografskih podataka radova [182]. Slika 16 predstavlja mrežu korelacije između ključnih reči, gde veličina čvora odgovara učestalosti pojavljivanja određene ključne reči, a debljina veza između ključnih reči odgovara relativnoj učestalosti zajedničkog navođenja između dve ključne reči. Boja čvorova i veza između njih odgovara srednjoj vrednosti godina u kojima su se određene ključne reči navodile, čija je legenda predstavljena u donjem desnom uglu mreže. Ključne reči su grupisane u 9 klastera, sa minimalnom veličinom klastera

postavljenom na 2. Ključna reč je morala biti navođena minimum 3 puta u okviru primarnih studija kako bi bila uključena u analizu. Klaster 1 je ujedno i najveći klaster ključnih reči i pozicioniran je na sredini mreže povezanosti. Čine ga ključne reči aopp, nadgledanje procesa, pomeranje procesa, poredjenje procesa, proizvodnja i analiza učenja. Klaster 2 sadrži 7 stavki, uključujući klastering, agregacija događaja, otkrivanje poslovnih procesa, predviđanje, simulacija i zdravstvena industrija. Klaster 3 sadrži 5 stavki, a to su analiza performansi, analiza uskih grla, *disco*, e-učenje i *moodle*. Klaster 4 sadrži 4 stavke, a to su internet stvari, mašinsko učenje, otkrivanja i analiza podataka i provera usaglašenosti. Klaster 5 sadrži 3 stavke, a to su dnevnik izvršenja događaja, fazi algoritam i otkrivanje poslovnog procesa. Klaster 6 sadrži 2 stavke *RPA* i upravljanje poslovnim procesima. Klaster 7 sadrži 2 stavke, a to su analiza ponašanja i *ProM*. Klaster 8 sadrži 2 stavke, a to su računarstvo u oblaku i nadgledanje. Klaster 9 sadrži 2 stavke, a to su digitalna transformacija i unapređenje procesa.

Može se zaključiti da su značajni klasteri koji sadrže više od 3 stavke. Na primer, klaster 1 predstavlja osnovne tipove automatskog otkrivanja poslovnih procesa koji se izvode u praksi. Klaster 2 pokazuje da se predviđanje, klastering i simulacija često izvode zajedno. Klaster 3 pokazuje da se u obrazovnoj industriji najčešće izvodi analiza performansi procesa, i to primenom fazi algoritma u latu *Disco*. Klaster 4 pokazuje da su internet stvari i mašinsko učenje povezani, kao da je provera usaglašenosti tip automatskog otkrivanja koji se uz njih najčešće primenjuje.

Takođe, sa same slike 16. moguće je uvideti da su najučestalije ključne reči aopp, zdravstvena industrija, otkrivanje poslovnih procesa, unapređenje poslovnih procesa i dnevnik izvršenja događaja.



Slika 16. Mreža povezanosti ključnih reči primarnih studija

3.1.4. Primenjene metode statističke obrade podataka

Nakon završetka prikupljanja podataka putem sistematskog pregleda literature, potrebno je odabrati odgovarajuće statističke metode radi testiranja postavljenih hipoteza u IM1. Odabir statističke metode zavisi od cilja analize, odnosno postavljenih istraživačkih pitanja i hipoteza, kao i od osobina analiziranih podataka [68,179]. Osnovne osobine podataka koje određuju odabir statističke metode su: tip podataka, merna skala podataka, veličina uzorka, broj varijabli i distribucija podataka. Takođe pored osobina podataka, same pretpostavke ili uslovi koje statističke metode postavljaju kako bi se koristile

mogu da ograniče odabir statističkih metoda. Sve statističke metode se mogu podeliti u dve grupe: parametrijske i neparametrijske [68,179]. Parametrijske metode koriste se za upoređivanje dve ili više grupa podataka, ali uz strogo postavljene pretpostavke. Specifične pretpostavke zavise od vrste statističke metoda koja se koristi, a one najčešće su: normalna distribucija podataka, nezavisnost podataka, homogenost podataka, podaci kontinuiranog tipa i postojanje linearne veze između zavisne i nezavisne varijable. Neparametrijske metode se koriste kada uzorak ne ispunjava pretpostavke odgovarajuće parametrijske metode, tako što za svaki neparametrijski test postoje odgovarajuće alternative.

Uzorak koji se analizira za potrebe IM1 je sačinjen od podataka prikupljenih sistematskim pregledom literature čije su osobine predstavljene u tabeli 7. Moguće je zaključiti da uzorak sadrži 13 kategoričkih varijabli, predstavljene putem nominalne merne skale, osim varijable godina, koja je kontinuirana. Uzorak sadrži 144 studije slučaja i nema normalnu raspodelu podataka. Varijabla Tip AOPP-a je višestruka varijabla koja sadrži odgovarajuće tipove automatskog otkrivanja poslovnih procesa.

Ukoliko se uzmu u obzir osobine posmatranog uzorka i pretpostavke za primenu parametrijskih statističkih metoda, zaključuje se da će za potrebe testiranja hipoteza IM1 biti primenjena statistička metoda iz neparametrijske grupe testova. Takođe, ukoliko se uzme u obzir da se hipoteze odnose na ispitivanje veze između posmatranih kategoričkih varijabli, zaključuje se da će biti primenjen Hi-kvadrat (χ^2) test nezavisnosti (engl. *Chi-square test of independence*)[68,179].

Hi-kvadrat test je neparametrijska metoda statističke analize koja se koristi prilikom ispitivanja verovatnoće nezavisnosti između dve kategoričke varijable. Prilikom primene Hi-kvadrat testa, nulta hipoteza pretpostavlja da ne postoji statistički značajna veza između posmatranih varijabli. Alternativna hipoteza pretpostavlja suprotno, odnosno da postoji statistički značajna veza između posmatranih varijabli. Računanjem statistike testa pod nazivom hi-kvadrat statistika i poređenjem te vrednosti sa kritičnom vrednosti dobijenom na osnovu stepena slobode (engl. *degree of freedom* - df) i željenog stepena značajnosti, test određuje da li su posmatrane razlike statistički značajne ili su slučajne pojave. Ukoliko je p-vrednost (engl. *p-value*) Hi-kvadrat testa manja od odabranog stepena značajnosti (npr., $p < 0.05$), smatra se da se nulta hipoteza odbacuje i zaključuje se da postoji statistički značajna veza između posmatranih varijabli.

Dodatno, moguće je odrediti jačinu veze dokazane pomoću Hi-kvadrat testa, računanjem Kramerovog V koeficijenta (engl. *Cramer's V coefficient*) [68]. Njegove vrednosti se kreću između 0 i 1, gde 0 ukazuje na nedostatak veze, a 1 ukazuje na savršenu vezu između varijabli. Prilikom tumačenja vrednosti koeficijenta, uzimaju se u obzir veličina uzorka i broj kategorija varijabli. Ukoliko se posmatra uzorak prikupljen sistematskim pregledom literature, vrednosti Kramerovog V koeficijenta se tumače na sledeći način [68]: slaba veza = 0.06, srednja = 0.17, jaka veza = 0.29.

Takođe, prilikom izvođenja Hi-kvadrat testa, kreira se tabela kontingencije (engl. *contingency table*). Tabela kontingencije sadrži kategorije jedne varijable u kolonama, a druge varijable u redovima. Čelije tabele kontingencije sadrže frekvencije ili procenat posmatranih kombinacija između kategorija varijabli i omogućavaju uvid u paterne i pravila povezanosti između varijabli.

Hi-kvadrat test takođe ima pretpostavke koje moraju biti ispunjenje kako bi rezultati bili pouzdani i validni. Prilikom računanja statistike Hi-kvadrat testa, računaju se i očekivane frekvencije poklapanja između kategorija unutar ćelija tabele kontingencije (engl. *expected cell frequency*). Pravilo koje se primenjuje vezano za očekivanu frekvenciju unutar ćelija je da najviše 20% posmatranih ćelija sme da ima očekivanu frekvenciju manju od 5. Problem Hi-kvadrat testa jeste taj što je zasnovan na aproksimaciji očekivanih frekvencija i nije pouzdan ukoliko nije ispunjena navedena pretpostavka ili ukoliko je uzorak previše mali. U suprotnom, potrebno je primeniti Fišerov egzaktni test (engl. *Fisher's exact test*).

Fišerov egzaktni test [183] je statistički test koji se koristi za određivanje značaja povezanosti između dve kategoričke varijable. Nulta hipoteza koja se postavlja je ta da ne postoji statistički značajna veza između posmatranih varijabli. Alternativna hipoteza je suprotna i tvrdi da pretpostavlja da postoji

statistički značajna veza između posmatranih varijabli. Fišerov egzaktni test se ne zasniva na aproksimaciji frekvencija ćelija tabele kontingencije, već računa verovatnoću povezanosti između varijabli primenom hipergeometrijske distribucije [183]. Stoga je p-vrednost koja je izračunata pomoću Fišerov egzaktnog testa uvek tačna vrednost. Ukoliko je p-vrednost (engl. *p-value*) Fišerov egzaktnog testa manja od odabranog stepena značajnosti (npr., $p < 0.05$), smatra se da se nulta hipoteza odbacuje i zaključuje se da postoji statistički značajna veza između posmatranih varijabli.

U skladu sa prethodno navedenim, tok primene statističkih metoda nad svakom hipotezom sadržanom u IM1 je sledeći:

- Priprema podataka za obradu numerisanjem kategoričkih varijabli u IBM SPSS alatu i kreiranje višestruke varijable Tip aopp;
- Analiza osnovnih karakteristika stavki (deskriptivna statističku analizu) koja predstavlja skup metoda koje daju opis rezultata i ima za cilj grupisanje, sređivanje i prikazivanje statističkih podataka, kao i određivanje osnovnih pokazatelja statističkih serija;
- Primena Hi-kvadrat testa radi testiranja hipoteza, uz kreiranje tabele kontingencije radi uvida u paterne odnosa između varijabli i računanje Kramerovog V koeficijenta radi određivanja jačine dokazane veze između varijabli;
- Ukoliko nisu ispunjene pretpostavke Hi-kvadrat testa, primena Fišerovog egzaktnog testa radi testiranja određene hipoteze.

Navedene statističke metode će biti primenjene u IBM SPSS statističkom alatu.

3.2. Metodologija prikupljanja i obrade podataka istraživačkog modela II

Istraživački model 2 sadrži hipotezu koja pretpostavlja postojanje veze između problema kvaliteta podataka dnevnika izvršenja događaja i tehnika čišćenja dnevnika izvršenja događaja. Metoda prikupljanja podataka za IM2 je upitnik, te je potrebno definisati dizajn upitnika. Dizajn upitnika se određuje kroz određivanje njegovih dimenzija, gde se merna dimenzija odnosi na definisanje konstrukata upitnika i njihovih tipova podataka, a reprezentaciona dimenzija na određivanje ciljne populacije upitnika [184]. Stoga su u ovom potpoglavlju predstavljeni razvoj i distribucija mernog instrumenta i njegovih konstrukata, opis ciljane populacije odnosno uzorka, statističke metode primenjene radi testiranja hipoteze i socio-demografska struktura ispitanika.

3.2.1. Razvoj i distribucija mernog instrumenta

Merna dimenzija upitnika se definiše kroz konstrukte mernog instrumenta. Tabela 6 sumira vrste problema kvaliteta podataka i grupe tehnika čišćenja dnevnika izvršenja događaja, definisanih na osnovu sistematskog pregleda literature i prethodnih istraživanja. Problemi kvaliteta dnevnika izvršenja događaja su izraženi kroz 22 različita problema, odnosno stavke upitnika, a tehnike čišćenja dnevnika izvršenja događaja su izražene kroz 7 grupa, odnosno stavki upitnika. Na osnovu prethodno navedenog i sa ciljem testiranja hipoteze IM2, kreiran je upitnik kao merni instrument. Upitnik je priložen kao Prilog B.

Potrebno je napomenuti da je Likertova skala najčešće korišćen tip psihometrijske rangirajuće skale koja ima sposobnost da meri stavove, mišljenja ili percepcije ispitanika na određenu temu [68,185]. Skala predstavljala niz tvrdnji ili stavki dimenzija upitnika, a ispitanici označavaju određeni intervalni nivo koji odgovara njihovoj percepciji. Na taj način, Likertova skala pruža struktuirani način kvantifikovanja kvalitativnih podataka i prikupljanja ljudskih mišljenja.

Uvodni deo upitnika objašnjava ispitanicima motivaciju i značaj istraživanja, i sadrži osnovne informacije o strukturi upitnika i vremenu potrebnom da se ispuni. Ispitanici su obavešteni da je upitnik anonimnog karaktera i da će se podaci koristiti isključivo u svrhe istraživanja predstavljene doktorske disertacije.

Druga sekcija – Demografija, sadrži značajna demografska pitanja koja se odnose na generalno iskustvo ispitanika u primeni tehnika čišćenja podataka. Prvo pitanje se odnosi na njihovo poznavanje

oblasti obrade podataka. Nivo poznavanja oblasti obrade podataka se meri pomoću petostepene Likertove skale, gde je:

1 – „slabo“, 2 – „umereno dobro“, 3 – „dobro“, 4 – „veoma dobro“ i 5 – „odlično“.

Drugo pitanje se odnosi na ulogu ispitanika u zajednici automatskog otkrivanja poslovnih procesa, gde kroz pitanje zatvorenog tipa, ispitanici imaju mogućnost da odaberu jedan od predefinisanih odgovora:

1 – „istraživač“, 2 – „praktikant“ ili 3 – „oboje“.

Zatim sledi pitanje zatvorenog tipa o vremenu istraživanja ili primene automatskog otkrivanja poslovnih procesa, gde ispitanici imaju mogućnost da odaberu vrednosti:

1 – „manje od jedne godine“, 2 – „1-5 godina“, 3 – „5-10 godina“ i 4 – „više od 10 godina“. Naredna su dva pitanja otvorenog tipa, gde se od ispitanika traži da navedu njihovo trenutno zanimanje i zemlju u kojoj su zaposleni. Poslednja dva pitanja se tiču softverskih alata koje najčešće primenjuju radi obrade podataka i radi primene ostalih tehnika automatskog otkrivanja poslovnih procesa. U oba pitanja, ispitanicima su ponuđeni najaktuelniji softverski alati kroz padajuću listu, ali imaju mogućnost dodavanja softverskog alata.

Treća sekcija – Problemi kvaliteta podataka, sadrži dve vrste pitanja koja se tiču stavki dimenzije problema kvaliteta podataka. Prvo pitanje se odnosi na značaj koji ispitanici pridaju konkretnim problemima kvaliteta podataka. Nivo percipiranog značaja problema kvaliteta meri se pomoću petostepene Likertove skale, gde je:

1 – „nije značajan“, 2 – „malo značajan“, 3 – „umereno značajan“, 4 – „značajan“ i 5 – „veoma značajan“.

Drugo pitanje se odnosi na učestalost sa kojom se ispitanici susreću sa konkretnim problemom kvaliteta u praksi. Nivo percipirane učestalosti problema kvaliteta meri se pomoću petostepene Likertove skale, gde je:

1 – „nikad“, 2 – „retko“, 3 – „povremeno“, 4 – „često“ i 5 – „veoma često“.

Četvrta sekcija – Tehnike čišćenja dnevnika izvršenja događaja, sadrži dve vrste pitanja koja se tiču stavki tehnika čišćenja dnevnika izvršenja događaja. Prvo pitanje se odnosi na značaj koji ispitanici pridaju konkretnim grupama tehnika čišćenja dnevnika izvršenja događaja. Nivo percipiranog značaja tehnika čišćenja dnevnika izvršenja događaja meri se pomoću petostepene Likertove skale, gde je:

1 – „nije značajna“, 2 – „malo značajna“, 3 – „umereno značajna“, 4 – „značajna“ i 5 – „veoma značajna“.

Drugo pitanje se odnosi na učestalost sa kojom ispitanici primenjuju konkretnu grupu tehnika čišćenja dnevnika izvršenja događaja kako bi umanjili ili uklonili problem kvaliteta u praksi. Nivo percipirane učestalosti primene tehnika čišćenja dnevnika izvršenja događaja meri se pomoću petostepene Likertove skale, gde je:

1 – „nikad“, 2 – „retko“, 3 – „povremeno“, 4 – „često“ i 5 – „veoma često“.

Poslednja sekcija – Odabir tehnika čišćenja dnevnika izvršenja događaja, sadrži 22 pitanja gde se od ispitanika traži da za svaki problem kvaliteta podataka odaberu jednu odgovarajuću grupu tehnika čišćenja dnevnika izvršenja događaja koja je, prema njihovom mišljenju, najpogodnija da umanj ili ukloni taj problem. Pitanja su struktuirana kao padajuće liste, sa ponuđenih 7 grupa tehnika čišćenja dnevnika izvršenja događaja za svaki problem i dodatnim poljem za unos drugih opcija, ukoliko ispitanik smatra da nijedna grupa tehnika nije pogodna za rešavanje konkretnog problema dnevnika izvršenja događaja.

3.2.2. Demografske karakteristike ispitanika

Reprezentaciona dimenzija istraživačkog dizajna IM2 se sastoji iz istraživača i praktikanata automatskog otkrivanja poslovnih procesa koji su upoznati sa problemima kvaliteta podataka i mogućim rešenjima. Uzorkovanje sa svrhom (engl. *purposive sampling*) je vrsta definisanja uzorka istraživanja kada postoji potreba da ciljani učesnici u istraživanju poseduju određene kvalitete, poput znanja ili iskustva u nekoj oblasti [186]. Uzorkovanje sa svrhom unapređuje rigoroznost istraživanja, verodostojnost prikupljenih podataka i dubinu razumevanja istraživanja od strane ispitanika [187,188]. Stoga je pristup uzorkovanja primenjen u ovom istraživanju uzorkovanje sa svrhom, sa uključivanjem totalne populacije koja ispunjava kriterijume.

Uzorak je sačinjen of članova IEEE radne grupe za automatsko otkrivanje poslovnih procesa, autora radova na temu problema kvaliteta podataka i njihovog uklanjanja i praktikanata automatskog otkrivanja poslovnih procesa sa trenutnim zanimanjem na *LinkedIn* platformi postavljenim na oblast automatskog otkrivanja poslovnih procesa. Upitnik je kreiran pomoću *SurveyMonkey* alata za internet upitnike i distribuiran je elektronskim putem [179], unapred definisanim redosledom. Poziv na učešće u istraživanju, zajedno sa linkom ka elektronskom upitniku je poslat svim potencijalnim učesnicima 15. januara 2023. godine. Podsetnici su slati u tri iteracije, u razmaku od nedelju dana. Upitnik je zatvoren 15. februara 2023. godine. Popunjavanje upitnika i učešće u istraživanju je bilo dobrovoljno.

Od ukupno 404 kontaktirana potencijalna učesnika u istraživanju, 230 je pristupilo elektronskom upitniku, dok je 207 popunilo kompletan upitnik. Stoga, može se zaključiti da je stopa odgovora bila 51.2%. kako bi se osigurao kvalitet rezultata istraživanja nastalih na osnovu obrade podataka prikupljenih upitnikom, učesnici koji su okarakterisali svoje iskustvo u oblasti obrade podataka „slabo“ su isključeni iz procedure obrade podataka. Na osnovu toga, konačan broj odgovora ispitanika koji je analiziran radi obrade podataka je 202.

Preostali ispitanici se na osnovu svog iskustva u obradi podataka mogu svrstati u 4 kategorije, predstavljene u tabeli 13. Najveći broj ispitanika je ocenio svoje iskustvo u obradi podataka kao „veoma dobro“, sa 36.8% od ukupnog broja ispitanika. Sledeće kategorije ispitanika jesu ispitanici koji su ocenili svoje iskustvo u obradi podataka kao „dobro“, sa učešćem od 29.9% i oni koji su ocenili svoje iskustvo u obradi podataka kao „odlično“, sa zastupljenošću od 23.1% od ukupnog uzorka. Najmanje zastupljena kategorija od ispitanika uključenih u istraživanje jesu oni ispitanici koji su ocenili svoje iskustvo u obradi podataka kao „umereno dobro“.

Tabela 13. Zastupljenost ispitanika uzorka prema iskustvu u obradi podataka

Iskustvo u obradi podataka	% ispitanika
Umereno dobro	10.2%
Dobro	29.9%
Veoma dobro	36.8%
Odlično	23.1%

Uloge ispitanika su u uzorku zastupljene skoro jednako, sa istraživačima koji obuhvataju 36.9% od ukupnog broja ispitanika i praktikanata koji obuhvataju 36.6% od ukupnog broja ispitanika. Treća kategorija uloga predstavlja ispitanike koji su se okarakterisali kao „oboje“ i čine 23.7% od ukupnog broja ispitanika u uzorku.

Ispitanici se prema iskustvu u primeni automatskog otkrivanja poslovnih procesa mogu podeliti u 4 grupe, predstavljene u tabeli 14. Naime, 57.4% ispitanika se bavi automatskim otkrivanjem poslovnih procesa između 1 i 5 godina, praćeni sa grupom koja se bavi automatskim otkrivanjem poslovnih procesa između 6 i 10 godina i koju činu 23.2% ispitanika. Zatim sledi grupa koja se bavi automatskim otkrivanjem poslovnih procesa više od 10 godina, čineći 13.8% od ukupnog broja ispitanika uzorka, praćena sa najmanje zastupljenom grupom sa 5.4% od ukupnog broja ispitanika uzorka koju čine ispitanici sa manje od 1 godine iskustva u automatskom otkrivanju poslovnih procesa.

Tabela 14. Zastupljenost ispitanika uzorka prema iskustvu u oblasti automatskog otkrivanja poslovnih procesa

Iskustvo u AOPP	% ispitanika
Manje od 1 godine	5.4%
1-5 godina	57.4%
6-10 godina	23.2%
Više od 10 godina	13.8%

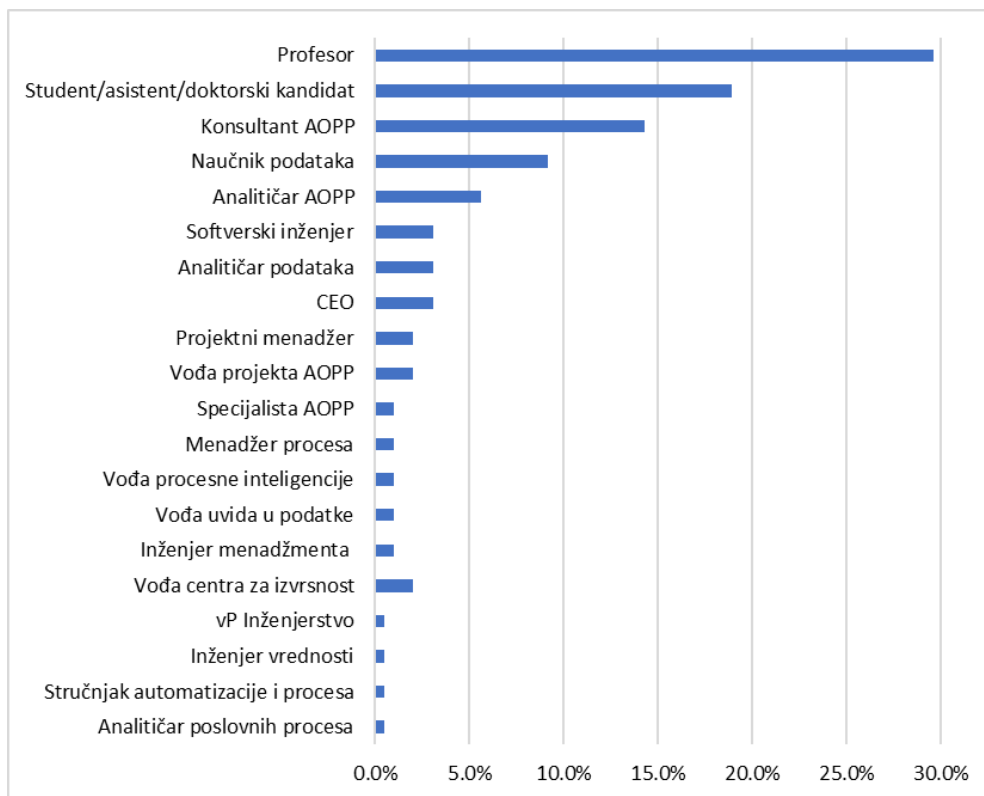
Kako bi se ispitale razlike između uloge koju ispitanici imaju i vremenskim periodom kojim se bave automatskim otkrivanjem poslovnih procesa, primenjen je Hi-hvadrat test nezavisnosti [68]. Rezultat pokazuje statistički značajan Hi-kvadrat test ($\chi^2 = 25.028$, $df 6$, $p = 0.000$), koji potvrđuje da ispitanici koji imaju različite uloge, takođe imaju različit vremenski period kojim se bave automatskim otkrivanjem poslovnih procesa. Tabela 15 prikazuje tabelu kontingencije koja je nastala prilikom sprovođenja Hi-kvadrat testa i koja pokazuje paterne razlike između ispitanika u pogledu uloga i vremenskih perioda kojim se bave automatskim otkrivanjem poslovnih procesa.

Tabela 15. Tabela kontingencije odnosa uloge i vremenskog perioda izučavanja/primene automatskog otkrivanja poslovnih procesa

		Vremenski period izučavanja/primene AOPP				Ukupno
		Manje od 1 godine	1-5 godina	6-10 godina	Više od 10 godina	
Uloge	Istraživač	2	45	15	18	80
	Praktikant	9	47	16	2	74
	Oba	0	24	16	8	48
Ukupno		11	116	47	28	202

Može se zaključiti da, iako većina ispitanika ima iskustvo u trajanju od 1 do 5 godina, istraživači imaju najviše ispitanika sa iskustvom dužim od 10 godina, dok praktikanti imaju u najvećoj meri iskustvo u trajanju od 1 do 5 godina.

Slika 17 predstavlja zastupljenost ispitanika uzorka prema zanimanju, sa 20 različitih kategorija zanimanja. Najčešće zanimanje ispitanika iz oblasti automatskog otkrivanja poslovnih procesa je zanimanje profesora, koje obuhvata 29.6% od ukupnog broja ispitanika. Naredno često zastupljeno zanimanje su takođe istraživačke pozicije, poput studenta, kandidata za doktorske studije ili asistenta, koji čine 18.9% od svih zanimanja ispitanika. Zatim, pojavljuje se zanimanje usko vezano za oblast, sa nazivom konsultant automatskog otkrivanja poslovnih procesa (engl. *process mining consultant*) koji čini 14.3% od svih zanimanja ispitanika. Naredno zanimanje kojim se ispitanici bave jeste naučnik podataka (engl. *data scientist*), koje obuhvata 9.2% od svih zanimanja ispitanika. Analitičar automatskog otkrivanja poslovnih procesa (engl. *process mining analyst*) zauzima 5.6% od svih zanimanja ispitanika. Ostala zanimanja, poput zanimanja generalnog direktora (engl. *Chief Executive Officer* – CEO) su zastupljena u manje od 5% slučajeva, a uvid u njih je moguć na slici 17.



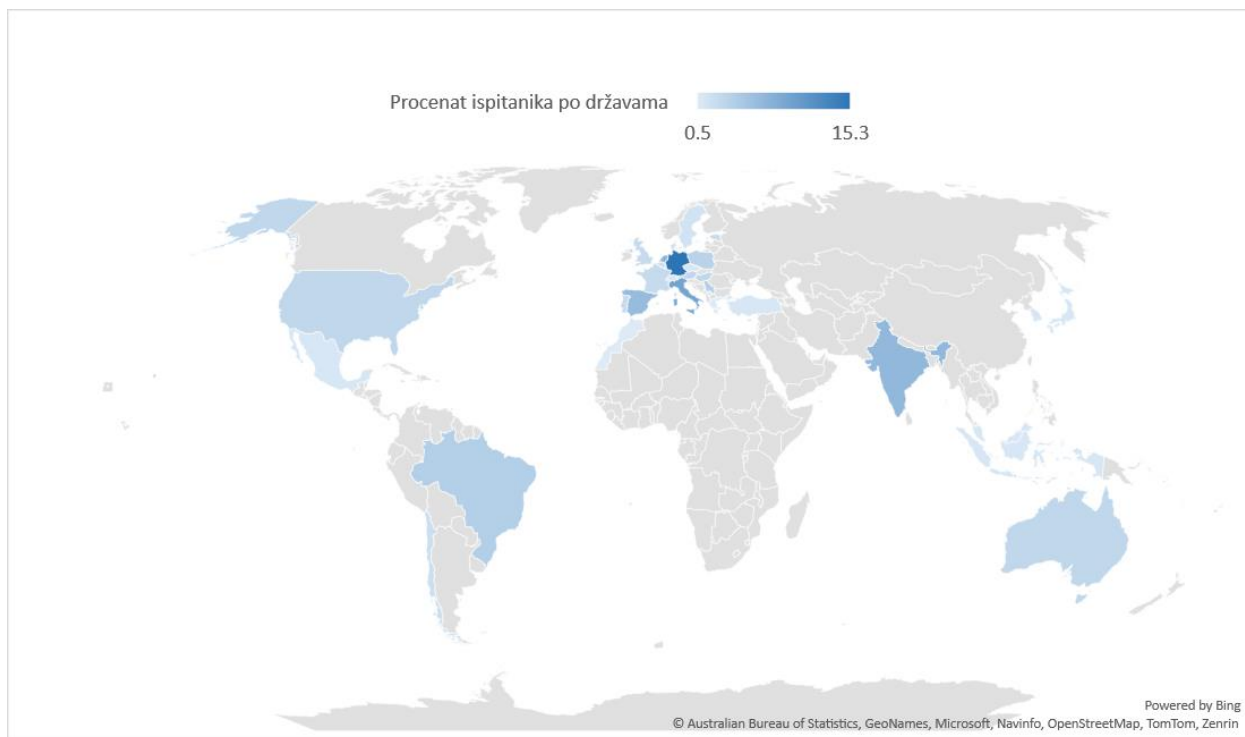
Slika 17. Zastupljenost ispitanika uzorka prema zanimanju

Prema tabeli 16 koja predstavlja kontinentalnu geografsku zastupljenost ispitanika uzorka, ispitanici su zaposleni u Evropi (71.0%), Aziji (15.5%), Južnoj Americi (7.0%), Australiji (3.0%), Sjedinjenim Američkim Državama (SAD) (3.0%) i Africi (0.5%).

Tabela 16. Zastupljenost ispitanika uzorka prema kontinentima na kojima su zaposleni

Kontinent	% ispitanika
Evropa	71.0%
Azija	15.5%
Južna Amerika	7.0%
Australija	3.0%
Sjedinjene Američke Države	3.0%
Afrika	0.5%

Slika 18 predstavlja mapu zastupljenosti država u kojima su ispitanici zaposleni, gde procenat zastupljenosti varirira od 0.5% do 15.3%. siva boja predstavlja države u kojima nema ispitanika, svetlo plava boja predstavlja države koje sadrže niži nivo ispitanika, a tamno plava boja označava države u kojima postoji veliki broj ispitanika. Konkretno, najveći procenat ispitanika uzorka je iz Nemačke, sa 15.3% od ukupnog broja ispitanika, praćeno sa Italijom, gde je zaposleno 9.4% od ukupnog broja ispitanika. Zatim slede Holandija, sa 7.4%, Indija sa 7.4%, Španija sa 6.4%, Brazil sa 4% i Poljska sa 3.5% od ukupnog broja ispitanika. Države u kojima je zaposleno 3.0% od ukupnog broja ispitanika su Australija, Austrija, Mađarska, Iran i SAD. Države u kojima je zaposleno 2.5% od ukupnog broja ispitanika su Portugal, Francuska, Estonija i Engleska. Zatim, države u kojima je zaposleno 2% od ukupnog broja ispitanika su Belgija, Čile i Danska. Države u kojima je zaposleno manje od 2% od ukupnog broja ispitanika su Švedska, Češka, Indonezija, Izrael, Japan, Lihtenštajn, Malezija, Meksiko, Južna Koreja, Meksiko, Švajcarska, Turska, Grčka i Maroko.



Slika 18. Mapa zastupljenosti ispitanika po državama

Zastupljenost ispitanika uzorka prema softverskom alatu za primenu automatskog otkrivanja poslovnih procesa, predstavljena u tabeli 17, pokazuje da je najzastupljeniji softverski alat za tu namenu *Celonis*, odabran od strane 27.7% od ukupnog broja ispitanika. Zatim sledi *ProM*, sa 20.3% udela, *Fluxicon Disco* sa 10.9% i *python* biblioteka za automatsko otkrivanje poslovnih procesa sa nazivom *PM4Py*, sa učešćem od 10.4% od ukupnog broja navedenih softverskih alata. Preostali softverski alati od značaja su *Apromore* (5%), *Noreja Process Intelligence* (3%), *SAP Signavio Process Intelligence* (3%), *BehfaLab* (2%), *R* (2%), *RapidProm* (1%), i drugi, sa učešćem manjim od 1% od ukupnog broja navedenih alata.

Tabela 17. Zastupljenost ispitanika uzorka prema softverskom alatu za primenu automatskog otkrivanja poslovnih procesa

Softverski alati	% ispitanika
<i>Celonis</i>	27.7%
<i>ProM</i>	20.3%
<i>Fluxicon Disco</i>	10.9%
<i>PM4Py</i>	10.4%
<i>Apromore</i>	5.0%
<i>Noreja Process Intelligence</i>	3.0%
<i>SAP Signavio Process Intelligence</i>	3.0%
<i>Behfa Lab</i>	2.0%
<i>R</i>	2.0%
<i>RapidProm</i>	1.0%
Drugi	<1%

Zastupljenost softverskih alata za obradu podataka u kontekstu automatskog otkrivanja poslovnih procesa je predstavljena u tabeli 18. Najzastupljeniji softverski alat za obradu podataka sa učešćem od 25.2% je *Celonis*. Zatim slede *PM4Py* sa učešćem od 16.8%, *ProM* sa učešćem od 15.8%, *Fluxicon Disco* sa učešćem od 5.0% i *R* sa učešćem od 5.0% od ukupnog broja korišćenih alata. Preostali alati su *R* (5.0%), *Structured Query Language – SQL* (4.5%), *SAP Signavio Process Intelligence* (3.0%),

Noreja Process Intelligence (3.0%), *Amazon Web Services – AWS* (3.0%), *Excel* (3.0%), *BehfaLab* (2.0%), *KNIME* (2.0%), *Pandas* (2%), *BupaR* (1.0%), i drugi, sa učešćem manjim od 1% od ukupnog broja navedenih alata.

Tabela 18. Zastupljenost ispitanika uzorka prema softverskom alatu za obradu dnevnika izvršenja događaja

Softverski alati	% ispitanika
<i>Celonis</i>	25.2%
<i>PM4Py</i>	16.8%
<i>ProM</i>	15.8%
<i>Fluxicon Disco</i>	5.0%
<i>R</i>	5.0%
<i>SQL</i>	4.5%
<i>Noreja Process Intelligence</i>	3.0%
<i>SAP Signavio Process Intelligence</i>	3.0%
<i>Excel</i>	3.0%
<i>AWS</i>	3.0%
<i>Pandas</i>	2.0%
<i>BehfaLab</i>	2.0%
<i>BupaR</i>	1.5%
Drugi	<1%

Kako bi ispitale razlike između uloge ispitanika i odabira softverskog alata za primenu automatskog otkrivanja poslovnih procesa, primenjen je Hi-kvadrat test nezavisnosti. Rezultati pokazuju statistički značajan Hi-kvadrat test ($\chi^2 = 179.925$, $df = 46$, $p = 0.000$), što potvrđuje da postoji veza između uloge ispitanika i njihove tendencija pri odabiru softverskog alata za primenu automatskog otkrivanja poslovnih procesa. Tabela 19 prikazuje tabelu kontingencije koja je nastala prilikom sprovođenja Hi-kvadrat testa i koja pokazuje paterne razlike između ispitanika u pogledu uloga i odabira softverskog alata za automatsko otkrivanje poslovnih procesa. Može se zaključiti da istraživači, više nego druge grupe, koriste alate *ProM*, *Fluxicon Disco* i *Aprimore*, dok praktikanti dominantno koriste *Celonis*. Oni koji ispitanici koji pripadaju istraživačima i praktikantima u najvećoj meri koriste *Celonis* i *ProM*.

Tabela 19. Tabela kontingencije odnosa uloge i softverskih alata za primenu automatskog otkrivanja poslovnih procesa

Softverski alati	Uloge		
	Istraživač	Praktikant	Oba
<i>Aprimore</i>	6	0	4
<i>Aris Process Mining</i>	0	0	2
<i>BehfaLab</i>	1	1	2
<i>bupaR</i>	2	0	0
<i>Celonis</i>	3	40	13
<i>Fluxicon Disco</i>	16	4	2
<i>icnMiner</i>	2	0	0
<i>iGrafx</i>	0	3	0
<i>Noreja Process Intelligence</i>	0	0	6
<i>PAFNow</i>	0	1	0
<i>PM4Py</i>	15	4	2
<i>ProM</i>	29	2	10
<i>QPR ProcessAnalyzer</i>	0	5	0

<i>R</i>	1	4	0
<i>RapidProm</i>	2	0	0
<i>SAP Signavio Process Intelligence</i>	0	5	1
<i>Tableau</i>	2	0	0
<i>Tibco Spotfire Template</i>	0	2	0
<i>UiPath</i>	0	0	4
<i>UPFLUX</i>	0	0	2

Kako bi ispitala razlike između uloge ispitanika i odabira softverskog alata za obradu podataka u kontekstu automatskog otkrivanja poslovnih procesa, primenjen je Hi-kvadrat test nezavisnosti (engl. *Chi-square test of independence*). Rezultati pokazuju statistički značajan Hi-kvadrat test ($\chi^2 = 180.805$, $df = 88$, $p = 0.000$), što potvrđuje da postoji veza između uloge ispitanika i njihove tendencija pri odabiru softverskog alata za obradu podataka pre izvođenja automatskog otkrivanja poslovnih procesa. Tabela 20 prikazuje tabelu kontingencije koja je nastala prilikom sprovođenja Hi-kvadrat testa i koja pokazuje obrasce razlike između ispitanika u pogledu uloga i odabira softverskog alata za obradu podataka pre izvođenja automatskog otkrivanja poslovnih procesa. Može se zaključiti da istraživači, više nego druge grupe, koriste alate *ProM*, *Fluxicon Disco* i *PM4Py*, dok praktikanti dominantno koriste *Celonis*, *R* i *SAP Signavio Process Intelligence*. Oni koji ispitanici koji pripadaju istraživačima i praktikantima u najvećoj meri koriste *Celonis* i *Noreja Process Intelligence* alate.

Tabela 20. Tabela kontingencije odnosa uloge i softverskih alata za obradu podataka u kontekstu automatskog otkrivanja poslovnih procesa

Softverski alati	Uloge		
	Istraživač	Praktikant	Oba
<i>Apromore</i>	0	0	2
<i>AWS</i>	0	8	0
<i>BehfaLab</i>	2	2	2
<i>bupaR</i>	3	0	0
<i>Celonis</i>	3	35	13
<i>Databricks with PySpark</i>	0	0	2
<i>Excel</i>	2	2	2
<i>Fluxicon Disco</i>	10	0	0
<i>KNIME</i>	0	4	0
<i>Noreja Process Intelligence</i>	0	0	6
<i>PAFNow</i>	0	1	0
<i>Pandas</i>	4	0	0
<i>PM4Py</i>	23	4	8
<i>ProM</i>	24	0	8
<i>QPR ProcessAnalyzer</i>	0	1	0
<i>R</i>	2	8	0
<i>Rapidminer</i>	0	2	0
<i>rho-Algorithm</i>	2	0	0
<i>SAP Signavio Process Intelligence</i>	0	5	1
<i>SQL</i>	5	2	2
<i>UPFLUX</i>	0	0	2

3.2.3. Primenjene metode statističke obrade podataka

Nakon završetka prikupljanja podataka putem upitnika, potrebno je odabrati odgovarajuće statističke metode radi testiranja postavljene hipoteze u IM2. Odabir statističke metode zavisi od cilja analize, odnosno postavljenih istraživačkih pitanja i hipoteza, kao i od osobina analiziranih podataka [68,179]. Osnovne osobine podataka koje određuju odabir statističke metode su: tip podataka, merna skala podataka, veličina uzorka, broj varijabli i distribucija podataka. Takođe pored osobina podataka, same pretpostavke ili uslovi koje statističke metode postavljaju kako bi se koristile mogu da ograniče odabir statističkih metoda.

Uzorak prikupljen putem upitnika sadrži 7 pitanja koja se tiču demografskih karakteristika ispitanika, čiji su podaci obrađeni primenom deskriptivne statistike, kako bi se dao opis rezultata i grupisano prikazali statistički podaci. Takođe, kako bi se ispitala razlike u pogledu uloga ispitanika i njihovog iskustva u primeni automatskog otkrivanja poslovnih procesa, kao i razlike u pogledu uloga ispitanika i njihovih preferencija u pogledu primene softverskih alata, primenjen je Hi-kvadrat test. Ispitane demografske karakteristike ispitanika su prikazane u potpoglavlju 3.2.2.

Po završetku procesa prikupljanja podataka putem upitnika, a s obzirom na prirodu podataka i postavljene hipoteze, pristupljeno je statističkoj obradi podataka koja uključuje:

- Deskriptivnu statističku analizu percipiranog značaja i učestalosti susretanja sa problemima kvaliteta dnevnika izvršenja događaja;
- Deskriptivnu statističku analizu percipiranog značaja i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja;
- Analizu značaja i učinka između percipiranog značaja i učestalosti susretanja sa problemima kvaliteta dnevnika izvršenja događaja;
- Analizu značaja i učinka između percipiranog značaja i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja;
- Primenu analize varijanse kako bi se ispitala razlike u značaju koji ispitanici različitih uloga pridaju problemima kvaliteta podataka dnevnika izvršenja događaja, kao i kako bi se ispitala razlike u učestalosti sa kojom ispitanici različitih uloga susreću probleme kvaliteta podataka dnevnika izvršenja događaja.
- Primenu analize varijanse kako bi se ispitala razlike u značaju koji ispitanici različitih uloga pridaju tehnikama čišćenja dnevnika izvršenja događaja, kao i kako bi se ispitala razlike u učestalosti sa kojom ispitanici različitih uloga primenjuju tehnike čišćenja dnevnika izvršenja događaja;
- Primenu Pirsonovog testa korelacije (engl. *Pearson's correlation test*) i Hi-kvadrat testa, kako bi ispitala hipoteza iz IM2. Pirsonov test korelacije se primenjuje kako bi se ispitalo da li postoji statistički značajna korelacija između učestalosti susretanja sa problemima kvaliteta podataka i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja. Hi-kvadrat test se primenjuje kako bi se ispitalo da li postoji statistički značajna veza između problema kvaliteta dnevnika i tehnika čišćenja dnevnika izvršenja događaja, na osnovu odabira tehnika čišćenja u zavisnosti od problema kvaliteta od strane ispitanika. Uz Hi-kvadrat test, kreira se i tabela kontingencije koja pokazuje paterne odabira tehnika čišćenja dnevnika izvršenja događaja na osnovu problema kvaliteta podataka.

Deskriptivna statistička analiza stavki upitnika podrazumeva računanje frekvencije ili udela koje određene stavke dimenzije problema kvaliteta i dimenzije tehnika čišćenja dnevnika izvršenja događaja imaju u pogledu percipiranog značaja i učestalosti primene, merenih na petostepenoj Likertovoj skali.

IPA analiza je zatim primenjena kako bi se stekao uvid u to koji problemi kvaliteta dnevnika podataka i tehnike čišćenja podataka treba da budu u fokusu istraživanja i zahtevaju posebnu pažnju, a koji nisu toliko značajni. IPA analiza ima mogućnost da utvrdi odnos između značaja koji ispitanici pridaju određenim stavkama i učestalosti sa kojom se ispitanici susreću sa određenim stavkama [189]. Postupak IPA analize podrazumeva postojanje dve različite skale merenja istih stavki, kao što su u ovom slučaju značaj i učestalost. Zatim se na osnovu odgovora ispitanika kreira dijagram rasprostranjenosti

(engl. *scatter plot*), koji kroz horizontalnu, x-osu predstavlja aritmetičku srednju vrednost značaja posmatranih stavki, a kroz vertikalnu, y-osu predstavlja aritmetičku srednju vrednost učestalosti pojavljivanja tih stavki. Za svaku stavku se računa srednja vrednost percipiranog značaja i učestalosti pojavljivanja, kako bi svaka tačka podataka na dijagramu rasprostranjenosti predstavljala srednju vrednost stavki u pogledu posmatranih dimenzija. X-osa i Y-osa se presecaju na medijani posmatranih dimenzija, čime se dijagram rasprostranjenosti, odnosno IPA matrica, deli na sledeća četiri kvadranta:

- Visoki percipirani značaj, visoka učestalost;
- Visoki percipirani značaj, niska učestalost;
- Niski percipirani značaj, visoka učestalost;
- Niski percipirani značaj, niska učestalost.

Tumačenje pozicija stavki unutar IPA matrice i pripadnost određenom kvadrantu se tumači različito u zavisnosti od posmatrane dimenzije istraživanja. Ukoliko se posmatra IPA matrica problema kvaliteta podataka dnevnika izvršenja događaja, oni problemi koji su procenjeni kao visoko značajni i učestali treba da budu u fokusu istraživanja, odnosno pripadaju kvadrantu „Visok prioritet”. Zatim, problemi kvaliteta podataka koji su procenjeni kao visoko značajni, ali retki, treba da nastave da se retko pojavljuju i pripadaju kvadrantu „Tako nastavi”. Problemi kvaliteta podataka koji su niskog značaja, ali se često susreću pripadaju kvadrantu „Srednji prioritet”. Problemi kvaliteta podataka koji su niskog značaja i ne susreću se često pripadaju kvadrantu „Nizak prioritet”.

Sa druge strane, ukoliko se posmatra IPA matrica tehnika čišćenja dnevnika izvršenja događaja, one tehnike koje su su procenjene kao visoko značajne i učestale pripadaju kvadrantu „Tako nastavi”. Zatim, tehnike koje su procenjene kao visoko značajne, ali su retko korišćene, treba da budu u fokusu istraživanja, odnosno pripadaju kvadrantu „Visok prioritet”. Tehnike čišćenja dnevnika izvršenja događaja koje su niskog značaja, ali se često primenjuju pripadaju kvadrantu „Srednji prioritet”. Tehnike čišćenja dnevnika izvršenja događaja koje su niskog značaja i ne primenjuju se često pripadaju kvadrantu „Nizak prioritet”.

Kako bi se ispitala razlike u značaju koji ispitanici različitih uloga pridaju problemima kvaliteta podataka dnevnika izvršenja događaja i tehnikama čišćenja dnevnika izvršenja događaja, kao i kako bi se ispitala razlike u učestalosti sa kojom ispitanici različitih uloga susreću probleme kvaliteta podataka dnevnika izvršenja događaja i primenjuju tehnike čišćenja dnevnika izvršenja događaja, primenjena je analiza varijanse. Jednosmerna ANOVA (engl. *one-way ANOVA*) [68] se koristi radi ispitivanja razlika između kategoričke nezavisne varijable i jedne kontinuirane zavisne varijable. Postavlja se nulta hipoteza da ne postoji statistički značajna razlika između grupa nezavisnih varijabli u pogledu srednjih vrednosti zavisne varijable. Jednosmerna ANOVA zatim računa značajnost testa, gde ukoliko je statistički značaj testa ($p < 0.05$), nulta hipoteza se odbacuje i pretpostavlja se da postoji statistički značajna razlika između grupa nezavisnih varijabli u pogledu srednjih vrednosti zavisne varijable. Takođe, ukoliko ANOVA test pokaže da postoji razlika između posmatranih varijabli, primenjuje se i post-hok *Least Significant Difference* – LSD test. LSD test poredi razlike između aritmetičkih srednjih vrednosti odgovora svake dve grupe ispitanika i daje informaciju o tome na koji način se grupe razlikuju.

U pogledu ovog istraživanja, uloge ispitanika se posmatraju kao nezavisne varijable, gde se ispituju razlike u značaju koji različite grupe ispitanika pridaju problemima kvaliteta podataka dnevnika izvršenja događaja i tehnikama čišćenja dnevnika izvršenja događaja, kao i razlike u učestalosti sa kojom ispitanici različitih uloga susreću probleme kvaliteta podataka dnevnika izvršenja događaja i primenjuju tehnike čišćenja dnevnika izvršenja događaja. Značaj i učestalost problema kvaliteta podataka su mereni na petostepenoj Likertovoj skali, te odgovaraju pretpostavkama koje jednosmerna ANOVA ima u pogledu zavisne varijable.

Ukoliko se uzmu u obzir osobine posmatranog uzorka i pretpostavke za primenu parametrijskih statističkih metoda, zaključuje se da će za potrebe testiranja hipoteze IM2 biti primenjena statistička metoda iz neparametrijske grupe testova. Takođe, ukoliko se uzme u obzir da se hipoteza odnosi na ispitivanje veze između posmatranih kategoričkih varijabli, zaključuje se da će biti primenjen Hi-kvadrat (χ^2) test nezavisnosti [68,179]. Takođe, na pitanja učestalosti susretanja sa problemima kvaliteta podataka i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja, koji su mereni

petostepenom Likertovom skalom, moguće je primeniti parametrijsku statističku metodu, odnosno Pirsonov test korelacije [68].

Hi-kvadrat test je neparametrijska metoda statističke analize koja se koristi prilikom ispitivanja verovatnoće nezavisnosti između dve kategoričke varijable. Postupak primene i tumačenje rezultata Hi-kvadrat testa je objašnjeno u potpoglavlju 3.1.4.

Dodatno, moguće je odrediti jačinu veze dokazane pomoću Hi-kvadrat testa, računanjem Kramerovog V koeficijenta [68], kao što je navedeno u potpoglavlju 3.1.4.

Takođe, prilikom izvođenja Hi-kvadrat testa, kreira se tabela kontingencije (engl. *contingency table*). Tabela kontingencije sadrži kategorije jedne varijable u kolonama, a druge varijable u redovima. Čelije tabele kontingencije sadrže frekvencije ili procenat posmatranih kombinacija između kategorija varijabli i omogućavaju uvid u paterne i pravila povezanosti između varijabli.

Hi-kvadrat test takođe ima pretpostavke koje moraju biti ispunjenje kako bi rezultati bili pouzdani i validni. Prilikom računanja statistike Hi-kvadrat testa, računaju se i očekivane frekvencije poklapanja između kategorija unutar ćelija tabele kontingencije (engl. *expected cell frequency*). Pravilo koje se primenjuje vezano za očekivanu frekvenciju unutar ćelija je da najviše 20% posmatranih ćelija sme da ima očekivanu frekvenciju manju od 5. Problem Hi-kvadrat testa jeste taj što je zasnovan na aproksimaciji očekivanih frekvencija i nije pouzdan ukoliko nije ispunjena navedena pretpostavka ili ukoliko je uzorak previše mali. U suprotnom, potrebno je primeniti Fišerov egzaktni test (engl. *Fisher's exact test*).

U pogledu hipoteze iz IM2, nulta hipoteza pretpostavlja da ne postoji statistički značajna veza između problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja.

Pirsonov test korelacije je primenjen kako bi se ispitalo da li postoji statistički značajna korelacija između učestalosti susretanja sa problemima kvaliteta podataka i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja. Prilikom izvođenja Pirsonovog testa korelacije, za svaku stavku se računa statistička značajnost „ p ”, kao i se korelacioni koeficijent „ r ”, koji opisuje jačinu i smer veze između stavki posmatranih dimenzija istraživanja. Rezultata se smatra statistički značajnim ukoliko je $p < 0.05$. Vrednosti korelacionog koeficijenta se, u kontekstu ovog istraživanja, mogu tumačiti na sledeći način [68]:

- $0.10 < r < 0.29$ – slaba pozitivna korelacija;
- $0.30 < r < 0.49$ – umerena pozitivna korelacija;
- $r > 0.50$ – jaka pozitivna korelacija.

Negativna vrednost korelacionog koeficijenta ukazuje na statističku vezu između dve varijable koja ima tendenciju promene u suprotnom smeru.

4. REZULTATI ISTRAŽIVANJA

Rezultati istraživanja su predstavljani kroz rezultate koji se tiču istraživačkog modela 1 i rezultate koji se tiču istraživačkog modela 2.

4.1. Rezultati istraživačkog modela I

Istraživački model 1 je imao za cilj da ispita 5 hipoteza koje se tiču postojanja povezanosti između različitih ključnih činioca projekata automatskog otkrivanja poslovnih procesa. Merni instrument korišćen za prikupljanje podataka je sistematski pregled literature studija slučaja u industrijskim sistemima. Naredna potpoglavlja predstavljaju rezultate primenjene metodologije obrade podataka nad prikupljenim podacima, kao što je navedeno u potpoglavlju 3.1.4. Prvo su predstavljeni rezultati deskriptivne statistike nad ključnim varijablama IM1, a zatim rezultati ispitivanja postavljenih hipoteza.

4.1.1. Deskriptivna statistika varijabli istraživačkog modela I

Ovo potpoglavlje predstavlja rezultate deskriptivne statistike varijabli istraživačkog modela 1, koje čine industrijski sistem, cilj analize, poslovni proces, tip automatskog otkrivanja poslovnih procesa, softverski alati i algoritmi i tehnike.

4.1.1.1. Industrijski sistem, cilj analize i poslovni procesi

Tabela 21 predstavlja deskriptivnu statistiku varijable industrija koja obuhvata učestalost pojavljivanja određene industrije u okviru posmatranih primarnih studija i procentualno izražen udeo određene industrije u odnosu na ukupan broj primarnih studija. Prilog A sadrži spisak svih primarnih studija sa referencama, podeljenim na osnovu industrije u kojima su sprovedene.

Tabela 21. Učestalost i udeo različitih vrsta industrijskih sistema

Industrija	Učestalost	Udeo u %
Zdravstvo	48	33.3
Proizvodnja	27	18.8
Obrazovanje	22	15.3
Informaciono-komunikacione tehnologije	20	13.9
Javni servisi i usluge	8	5.6
Računovodstvo, kontroling, bankarstvo i osiguranje	7	4.9
Logistika	6	4.2
Građevina	5	3.5
Energija	1	0.7
Ukupno	144	100.0

Tabela 22 predstavlja deskriptivnu statistiku varijable cilj analize koja obuhvata učestalost pojavljivanja određenog cilja analize u okviru posmatranih primarnih studija i procentualno izražen udeo određenog cilja analize u odnosu na ukupan broj primarnih studija. Prilog A sadrži spisak svih primarnih studija sa referencama, podeljenim na osnovu ciljeva analize.

Tabela 22. Učestalost i udeo različitih ciljeva analize

Cilj analize	Učestalost	Udeo u %
Otkrivanje i unapređenje poslovnog procesa	95	66.0
Predviđanje ponašanja procesa	13	9.0

Pronalazak anomalija i nepredviđenih aktivnosti	13	9.0
Poređenje poslovnih procesa	6	4.2
Analiza vremena čekanja	5	3.5
Analiza promena procesa	3	2.1
Automatizacija poslovnog procesa	3	2.1
Analiza resursa	3	2.1
Otkrivanje uzroka problema	2	1.4
Otkrivanje poslovnih pravila	1	0.7
Ukupno	144	100.0

Tabela 23 predstavlja deskriptivnu statistiku varijable poslovni proces koja obuhvata učestalost pojavljivanja određenog poslovnog procesa u okviru posmatranih primarnih studija i procentualno izražen udeo određenog poslovnog procesa u odnosu na ukupan broj primarnih studija. Prilog A sadrži spisak svih primarnih studija sa referencama, podeljenim na osnovu industrije, cilja analize i poslovnih procesa.

Tabela 23. Učestalost i udeo različitih poslovnih procesa

Poslovni proces	Učestalost	Udeo u %
Putanja pacijenta	30	20.8
Ponašanje korisnika	15	10.4
Proizvodni proces	13	9.0
Proces učenja	11	7.6
Proces izrade softvera	7	4.9
Upravljanje nepredviđenim okolnostima	6	4.2
Nabavka	6	4.2
Proces odgovora hitne službe	6	4.2
Prodaja	5	3.5
Izgradnja	4	2.8
Ocenjivanje studenata	3	2.1
Proces prijave	3	2.1
Proces sajber napada	3	2.1
Montaža	2	1.4
Obrada žalbe	2	1.4
Upravljanje resursima	2	1.4
Kreiranje filogeneze virusa	1	0.7
Laboratorijsko testiranje	1	0.7
Nadgledanje krvnog pritiska	1	0.7
Operacija	1	0.7
Post-operativna nega	1	0.7
Kretanje robe	1	0.7
Transport pacijenata	1	0.7
Priprema nastave	1	0.7
Upravljanje bezbednosnim rizikom	1	0.7
Obrada tužbe	1	0.7
Putanja studenta	1	0.7
Procena troškova	1	0.7

Naplata medicinskih računa	1	0.7
Grupno rešavanje problema	1	0.7
Prijava za kredit	1	0.7
Tok bolesti	1	0.7
Konfiguracija sistema	1	0.7
Pregled godišnjeg bilansa računa	1	0.7
Proces popravke greške pri radu softvera	1	0.7
Interna kontrola	1	0.7
Upravljanje projektima	1	0.7
Upravljanje skladištem	1	0.7
Proces preuzimanja prtljaga	1	0.7
Planiranje proizvodnje	1	0.7
Servisiranje automobila	1	0.7
Topologija fabrike	1	0.7
Ukupno	144	100.0

Zdravstvena industrija ima najveću učestalost u odnosu na ukupan broj posmatranih primarnih studija, sa 33.3% udela. Primarna studija pripada zdravstvenoj industriji ukoliko je predmet analize poslovni proces koji se odvija unutar sistema zdravstvene zaštite. Prilog A sadrži navedene sve poslovne procese koji su analizirani u okviru zdravstvene industrije, sa navedenim ciljevima analize.

Najčešće analiziran scenario u okviru zdravstvene industrije podrazumeva otkrivanje i unapređenje putanje pacijenta (engl. *patient pathway*) kroz zdravstveni sistem [30,190–206], koja se definiše na osnovu podataka dobijenih iz zdravstvenog sistema bolnice (engl. *Hospital Information System*). Neke primarne studije zdravstvene industrije su težile poređenju poslovnih procesa putanja pacijenata sa različitim karakteristikama iste bolesti [207–209] i poređenju različitih pristupa post-operativne nege [210]. Putanja pacijenta je analizirana iz perspektive analize vremena čekanja između aktivnosti procesa [211,212] i sa ciljem predviđanja toka kretanja pacijenta kroz zdravstveni sistem [213–216]. Neki autori su analizirali putanju pacijenta u kontekstu promena koje proces prolazi tokom samog izvršavanja [217], kao i radi planiranja resursa na osnovu toka kretanja pacijenata unutar zdravstvenog sistema [218,219]. Otkrivanje modela ponašanja subjekata analize u okviru zdravstvenih sistema je takođe bio čest scenario primene automatskog otkrivanja poslovnih procesa u zdravstvenoj industriji [220–223]. Autori su težili i otkrivanju i unapređenju modela procesa odgovora hitne službe [224–227]. Ostali poslovni procesi analizirani u okviru zdravstvene industrije su naplata medicinskih računa [228], tok bolesti [229], proces izvršavanja operacije [230], transport pacijenata [231], laboratorijsko testiranje [232], proces prodaje [233] i nadgledanje krvnog pritiska [234].

Proizvodna industrija ima udeo od 18.8% u odnosu na ukupan broj posmatranih primarnih studija. Primarna studija pripada proizvodnoj industriji ukoliko je predmet analize poslovni proces koji se odvija unutar proizvodnog sistema neke organizacije, sa ciljem izvršavanja osnovne funkcije tog sistema. Prilog A sadrži navedene sve poslovne procese koji su analizirani u okviru proizvodne industrije, sa navedenim ciljevima analize.

Najčešće analiziran scenario u okviru proizvodne industrije jeste otkrivanje modela i unapređenje procesa izvršavanja proizvodnje [235–237], kao i niz dodatnih pristupa analizi proizvodnog procesa. Dodatni ciljevi prilikom analize proizvodnog procesa su pronalazak anomalija i nepredviđenih okolnosti u izvršavanju proizvodnog procesa [71], analiza promena procesa koje se javljaju tokom izvršavanja proizvodnje [70], predviđanje ponašanja proizvodnog procesa [96,97,238], analiza vremena čekanja između aktivnosti proizvodnog procesa [239,240] i otkrivanje postojanja poslovnih pravila u okviru proizvodnog procesa [241]. Takođe značajni poslovni procesi u okviru proizvodnog sistema su procesi nabavke [85,242,243] i prodaje [90,101,244]. Upravljanje nepredviđenim okolnostima je poslovni proces unutar proizvodnog sistema koji ima za cilj da predvidi i umanju posledice nepredviđenih okolnosti. Autori su analizirali ovaj proces sa ciljem otkrivanja modela izvršavanja procesa i njegovog

unapređenja [95,245–247]. Autori su takođe primenjivali automatsko otkrivanje poslovnih procesa kako bi otkrili model proces planiranja proizvodnje [248] i model toplotologije fabrike [102]. Još jedna primena automatskog otkrivanja poslovnih procesa je bila analiza vremena čekanja i uskih grla pilikom izvođenja procesa montaže gotovih proizvoda [249]. Dalje, autori su uspjeli da automatizuju proces nabavke na osnovu modela procesa dobijenog pomoću tehnika otkrivanja procesa [250]. Konačno, proces upravljanja promenama i kontinuiranim unapređenjem je analiziran sa ciljem pronalaska anomalija i nepredviđenih okolnosti [251].

Obrazovna industrija ima udeo od 15.3% u odnosu na ukupan broj primarnih studija. Primarna studija pripada obrazovnoj industriji ukoliko je predmet analize poslovni proces koji se odvija unutar obrazovne institucije ili se odnosi na poslovni proces učenja i ocenjivanja. Prilog A sadrži navedene sve poslovne procese koji su analizirani u okviru obrazovne industrije, sa navedenim ciljevima analize.

Najčešće analiziran scenario u okviru obrazovne industrije jeste otkrivanje i unapređenje procesa učenja [252–259]. Proces učenja podrazumeva aktivnosti koje student izvršavaju na platformama elektronskog učenja tokom pohađanja kurseva. Autori su uspešno predvideli konačnu ocenu studenta na osnovu analize procesa učenja [260] i otkrili uzroke problema prilikom učenja studenata [261]. Takođe, analizirali su interakciju korsnika na platformama za elektronsko učenje kako bi automatizovali transfer podataka između studenata [262] i pronašli su anomalije u procesu upravljanja resursima platforma za elektronsko učenje [263]. Takođe, izvršeno je poređenje procesa ocenjivanja studenata na kosnovu odabranih kriterijuma [264]. Ostali poslovni procesi analizirani sa ciljem otkrivanja i unapređenja modela procesa su proces ocenjivanja studenata [169,265], upravljanje obrazovnim projektima [266], priprema nastave [267], grupno rešavanje problema [268], proces izrade softvera [269], upravljanje resursima [270] i ponašanje studenata [271].

IKT industrija ima udeo od 13.9% u odnosu na ukupan broj primarnih studija. Primarna studija pripada IKT industriji ukoliko je predmet analize poslovni proces koji se odvija unutar nekog softverskog sistema ili pametnog uređaja, proces izrade softvera ili analiza rada softvera. Prilog A sadrži navedene sve poslovne procese koji su analizirani u okviru obrazovne industrije, sa navedenim ciljevima analize.

Najčešće analiziran scenario u okviru IKT industrije jeste otkrivanje i unapređenje modela procesa razvoja softvera [272,273], kao i pronalazak anomalija prilikom procesa razvoja softvera [274–276]. Dalje, predmet analize u IKT industriji je često i modelovanje ponašanja korisnika nekog informacionog sistema [277–280], zatim predviđanje ponašanja korisnika nekog informacionog sistema [281] i pronalazak anomalija ponašanja korisnika [282]. Proces sajber napada je takođe analiziran kaok bi se otkrio model njegovog izvršavanja [171,283] i kako bi se utvrdile anomalije u ponašanju tokom sajber napada [284]. Poslovni procesi koji takođe analizirani u IKT industriji su proces popravke greške u radu softvera [285], kreiranje filogeneze virusa [286], upravljanje nepredviđenim okolnostima [287] i konfiguracija sistema [288].

Industrija javnih servisa i usluga ima udeo od 5.6% u odnosu na ukupan broj primarnih studija. Primarna studija pripada industriji javnih servisa i usluga ukoliko je predmet analize poslovni proces koji predstavlja javni servis ili uslugu koja se pruža građanima od strane državnih organa ili privatnih institucija. Prilog A sadrži navedene sve poslovne procese koji su analizirani u okviru industrije javnih servisa i usluga, sa navedenim ciljevima analize.

Najčešće analizirani poslovni procesi u okviru industrije javnih servisa i usluga su procesi podnošenja prijave za uslugu [289,290], koji su analizirani radi pronalaska anomalija ili otkrivanja uzroka problema procesa, kao i otkrivanje modela ponašanja korisnika prilikom korišćenja usluge [291–293]. Ostali procesi čiji su modeli procesi otkriveni i unapređeni su proces servisiranja automobila [294] i proces obrade sudske tužbe [295].

Industrija računovodstva, kontrolinga, bankarstva i osiguranja ima udeo od 4.9% u odnosu na ukupan broj primarnih studija. Primarna studija pripada industriji računovodstva, kontrolinga, bankarstva i osiguranja ukoliko je predmet analize poslovni proces koji proizilazi iz osnovnih aktivnosti

navedenih oblasti. Prilog A sadrži navedene sve poslovne procese koji su analizirani u okviru industrije računovodstva, kontrolinga, bankarstva i osiguranja, sa navedenim ciljevima analize.

Prijava za kredit i obrada žalbe su dva najčešće analizirana procesa ove industrije. Autori su uspešno otkrili i unapredili proces prijave za kredit [296], kao i proces obrade žalbe [297]. Dalje, primenili su prediktivno automatsko otkrivanje poslovnih procesa nad procesom prijave za kredit, kako bi ga automatizovali [298] i poredili su različito izvođenje obrade žalbe iz oblasti osiguranja kako bi definisali najbolje prakse [299]. Nad procesom pregleda godišnjeg bilansa računa primenili su pristup pronalaska anomalija i nepredviđenih aktivnosti, kako bi se uočila odstupanja u plaćanju [300]. Takođe, primenili su rezultate primene automatskog otkrivanja poslovnih procesa na proces interne kontrole, kako bi ga automatizovali [170]. Konačno, otkrili su model procesa kontrolinga [301].

Industrija logistike ima udeo od 4.2% u odnosu na ukupan broj primarnih studija. Primarna studija pripada industriji logistike ukoliko je predmet analize poslovni proces koji se bavi nabavkom, distribucijom i skladištenjem robe u okviru lanca snabdevanja ili pružanja javnih usluga. Prilog A sadrži navedene sve poslovne procese koji su analizirani u okviru industrije logistike, sa navedenim ciljevima analize.

Otkrivanje model procesa i njegovo unapređenje je u oblasti logistike sprovedeno nad procesima nabavke [302], prodaje [303] i upravljanja skladištem [304]. Interna logistika je analizirana u sklopu analize resursa koji se koriste prilikom procesa montaže [88]. Zatim, u okviru operativne logistike, uspešno je predviđen tok preuzimanja prtljaga na aerodromu [305]. Konačno, analizirane su promene u procesu kretanja robe koje se javljaju tokom samog izvršavanja procesa [306].

Građevinska industrija ima udeo od 3.5% u odnosu na ukupan broj primarnih studija. Primarna studija pripada građevinskoj industriji ukoliko je predmet analize poslovni proces koji se tiče planiranja i izvođenja gradnje i povezanih poslova. Prilog A sadrži navedene sve poslovne procese koji su analizirani u okviru građevinske industrije, sa navedenim ciljevima analize.

Proces izgradnje je analiziran sa nekoliko različitih ciljeva analize. U najvećoj meri, cilj analize je bio otkrivanje i unapređenje izvođenja procesa gradnje [307–309]. Zatim, uspešno je izvršeno predviđanje procesa gradnje u oblasti teške civilne gradnje [310]. Konačno, analiziran je i proces procene troškova građevinske kompanije [311].

Energetska industrija zauzima 0.7% od svih posmatranih primarnih studija. Njoj pripada jedna primarna studija, koja predstavlja rezultate primene automatskog otkrivanja poslovnih procesa nad procesom upravljanja bezbednosnim rezicima u energetske kompaniji [312].

4.1.1.2. Tipovi automatskog otkrivanja poslovnih procesa

Tabela 24 predstavlja učestalost i udeo različitih tipova automatskog otkrivanja poslovnih procesa koji su primenjeni u primarnim studijama. Jedna primarna studija je mogla da primeni jedan ili više tipova automatskog otkrivanja poslovnih procesa. Tip sa najvećim udelom primene jeste otkrivanje poslovnog procesa, sa udelom od 95.1%. Zatim sledi unapređenje poslovnog procesa, sa udelom od 76.4%. Značajno manju primenu imaju tipovi proveravanja usaglašenosti, sa 18.7% udela, analiza društvenih mreža, sa 9% udela i prediktivno automatsko otkrivanje poslovnih procesa, sa udelom od 5.5%.

Tabela 24. Učestalost i udeo različitih tipova automatskog otkrivanja poslovnih procesa

Tip AOPP	Učestalost	Udeo u %
Otkrivanje poslovnog procesa	137	95.1
Unapređenje poslovnog procesa	110	76.4
Provera usaglašenosti	27	18.7
Analiza društvenih mreža	13	9.0
Prediktivno AOPP	8	5.5
Ukupno	144	/

4.1.1.3. Algoritmi, tehnike i softverski alati

Tabela 25 predstavlja učestalost i udeo različitih algoritama za otkrivanje poslovnih procesa. Posmatrani broj primarnih studija je 137, jer je u toliko slučajeva izvršen tip otkrivanje poslovnog procesa. Tabela 25 sadrži 28 različitih algoritama.

Prominenti algoritmi za otkrivanje poslovnih procesa su fazi algoritam, sa udelom od 32.8%, induktivni algoritam, sa udelom od 24.8% i heuristički algoritam, sa udelom od 11.7%. Nakon njih sledi *Celonis* algoritam za otkrivanje modela procesa (udeo 5.8%), koji predstavlja algoritam za otkrivanje poslovnih procesa koji je ugrađen u softverski alat *Celonis*. Generator procesne mape je algoritam za otkrivanje modela procesa koji je ugrađen u softverski alat *BupaR* i primenjen je u 2.9% slučajeva primene otkrivanja poslovnih procesa. Interaktivni heuristički algoritam, primenjen u 2.9% slučajeva, je proširena varijanta heurističkog algoritma koji pruža mogućnost promene pogleda na model procesa. Alfa algoritam je primenjen u 2.2% posmatranih primarnih studija, ka oi algoritam za otkrivanje modela procesa ugrađen u *PALIA* softverski alat.

Preostali algoritmi za otkrivanje modela procesa su retko primenjivani algoritmi specifične namene, primenjeni u 0.7% posmatranih primarnih studija.

Tabela 25. Učestalost i udeo različitih algoritama za otkrivanje poslovnih procesa

Algoritam za otkrivanje poslovnih procesa	Učestalost	Udeo u %
Fazi algoritam	45	32.8
Induktivni algoritam	34	24.8
Heuristički algoritam	16	11.7
<i>Celonis</i> algoritam otkrivanja procesa	8	5.8
Generator procesne mape	4	2.9
Interaktivni heuristički algoritam	4	2.9
Alfa algoritam	3	2.2
<i>PALIA</i> algoritam otkrivanja procesa	3	2.2
Algoritam za otkrivanje frekvencija	1	0.7
Algoritam za otkrivanje promena procesa	1	0.7
Algoritam kauzalnih mreža otkrivanja procesa	1	0.7
<i>CloFast</i> sekvencijalni algoritam	1	0.7
Algoritam za distribuirano i deklarativno otkrivanje procesa	1	0.7
<i>ProM</i> dodatak za proveru usaglašenosti	1	0.7
<i>ProM</i> dodatak za otkrivanje procesa hitne službe	1	0.7
<i>EverFlow</i> algoritam otkrivanja procesa	1	0.7
Markov model prvog reda	1	0.7
Stablo za otkrivanje hibridnih aktivnosti	1	0.7
Vizualni induktivni algoritam	1	0.7
Algoritam za interaktivno otkrivanje poslovnih procesa	1	0.7
Algoritam za otkrivanje lokalnih modela procesa	1	0.7
LTL proverivač*	1	0.7
<i>Minit</i> algoritam otkrivanja procesa	1	0.7
<i>PAFnow</i> algoritam otkrivanja procesa	1	0.7
<i>process.analyser</i> algoritam otkrivanja procesa	1	0.7

<i>ProDiGen</i> algoritam otkrivanja procesa	1	0.7
<i>ProFit</i> pajton biblioteka	1	0.7
<i>WoMen</i> algoritam otkrivanja procesa	1	0.7
Ukupno	137	100.0

*Linear Temporal Logic – LTL

Tabela 26 predstavlja učestalost i udeo softverskih alata za otkrivanje poslovnih procesa. Najčešće primenjivani softverski alat za otkrivanje poslovnih procesa je *ProM*, sa udelom od 40.1% od posmatranih primarnih studija. *ProM* ima mogućnost primene fazi algoritma, heurističkog algoritma i induktivnog algoritma. Softverski alat *Disco*, sa mogućnošću primene samo fazi algoritma, ima udeo od 28.5% od posmatranih primarnih studija. Naredni softverski alat za otkrivanje poslovnih procesa je *Celonis*, sa udelom od 5.8% i ugrađenim algoritmom za otkrivanje poslovnih procesa. U 4.4% slučajeva, autori nisu koristili postojeći softverski alat, već su razvili sopstveni alat koji podržava njihov pristup analizi. Zatim, u 3.6% slučajeva autori su koristili pajton biblioteku za automatsko otkrivanje poslovnih procesa – *PM4Py*, u 2.9% slučajeva *BupaR* i u 2.2% slučajeva softverski alat *PALIA*. Preostali softverski alati za otkrivanje poslovnih procesa su korišćeni u pojedinačnim slučajevima.

Tabela 26. Učestalost i udeo softverskih alata za otkrivanje poslovnih procesa

Softverski alat za otkrivanje poslovnih procesa	Učestalost	Udeo u %
<i>ProM</i>	55	40.1
<i>Disco</i>	39	28.5
<i>Celonis</i>	8	5.8
Alat razvijen od strane autora primarne studije	6	4.4
<i>BupaR</i>	4	2.9
<i>PM4Py</i>	5	3.6
<i>PALIA</i>	3	2.2
Alat za interaktivno otkrivanje poslovnih procesa	1	0.7
<i>Apromore</i>	1	0.7
<i>Declare</i>	1	0.7
<i>EverFlow</i>	1	0.7
<i>MD - Declare</i>	1	0.7
<i>Minit</i>	1	0.7
<i>Networkx2</i>	1	0.7
<i>PAFnow</i>	1	0.7
<i>pMineR</i>	1	0.7
<i>process.analyser</i>	1	0.7
<i>ProDiGen</i>	1	0.7
<i>Rio Diagnostics</i>	1	0.7
<i>SOWCompact</i>	1	0.7
<i>Upflux</i>	1	0.7
<i>WoMen</i>	1	0.7
Ukupno	137	100.0

Tabela 27 predstavlja učestalost i udeo tehnika za predviđanje u automatskom otkrivanju poslovnih procesa. Može se zaključiti da se u najvećem broju slučajeva u primarnim studijama kao tehnika za predviđanje ponašanja procesa navodio neki od algoritama mašinskog učenja (62.5%). Međutim, u nekim primarnim studijama autori su navodili i specifične tehnike, poput bajezovih mreža, diskretne simulacije događaja i LSTM neuronskih mreža.

Tabela 27. Učestalost i udeo tehnika predviđanja u automatskom otkrivanju poslovnih procesa

Tehnike predviđanja	Učestalost	Udeo u %
Algoritam mašinskog učenja	5	62.5
Bajezove mreže	1	12.5
Diskretna simulacija događaja	1	12.5
LSTM neuronske mreže	1	12.5
Ukupno	8	100.0

Tabela 28 predstavlja učestalost i udeo primenjenih tehnika za unapređenje poslovnih procesa. Ukupan broj posmatranih primarnih studija odgovara broju primarnih studija koje su primenile tip unapređenje poslovnih procesa. Može se zaključiti da su najčešće primenjene tehnike unapređenja poslovnih procesa ona koje su ugrađene u algoritam za otkrivanje poslovnih procesa. Takve tehnike unapređenja prilikom definisanja modela procesa računaju metrike unapređenja, poput vremena trajanja aktivnosti, vremena čekanja između aktivnosti i ukazuju na uska grla u izvršavanju procesa. Vodeći algoritam za unapređenje poslovnih procesa je fazi algoritam (35.5%), praćen sa heurističkim algoritmom (12.7%), induktivnim algoritmom (12.7%) i ugrađenim tehnikama unapređenja u softverskom alatu *Celonis* (6.4%). *ProM* dodatak za proveru usaglašenosti između modela procesa i dnevnika izvršenja događaja je takođe korišćen u 3.6% posmatranih primarnih studija, kako bi se pronašle nelogičnosti u izvršavanju procesa. Generator procesne mape sa merama performansi, primenjen u 3.6% posmatranih primarnih studija, jeste proširenje algoritma za otkrivanje poslovnih procesa softverskog alata *BupaR*, koji osim modela procesa generiše i podatke o performansama procesa, odnosno podatke o vremenu trajanja izvršavanja aktivnosti i vremenu čekanja između njih. Analiza uskih grla je primenjena kao standardna metoda unapređenja poslovnih procesa u 2.7% posmatranih primarnih studija, sa ciljem pronalaženja aktivnosti procesa kod kojih postoji zastoj prilikom izvođenja. Softverski alat *PALIA* takođe ima mogućnost primene tehnika za unapređenje poslovnih procesa i primenjen je u 2.7% primarnih studija. Analiza paterna je primenjena u 1.8% primarnih studija i podrazumeva analizu redosleda izvršavanja između aktivnosti radi prepoznavanja paterna ponašanja unutar procesa. *ProM* dodatak za predikciju je takođe korišćen u 1.8% posmatranih primarnih studija, kako bi se predvidelo vreme trajanja izvršavanja procesa, na osnovu istorijskih podataka zabeleženih u dnevniku izvršavanja događaja. Interaktivni heuristički algoritam je primenjen u 1.8% primarnih studija, sa ugrađenim tehnikama koje nad modelom procesa prikazuju i mere performansi modela procesa. Preostale tehnike unapređenja poslovnih procesa su primenjene u pojedinačnim slučajevima.

Tabela 28. Učestalost i udeo tehnika za unapređenje poslovnih procesa

Tehnike za unapređenje poslovnih procesa	Učestalost	Udeo u %
Fazi algoritam – ugrađene tehnike unapređenja	39	35.5
Heuristički algoritam – ugrađene tehnike unapređenja	14	12.7
Induktivni algoritam – ugrađene tehnike unapređenja	14	12.7
<i>Celonis</i> – ugrađene tehnike unapređenja	7	6.4
<i>ProM</i> dodatak za proveru usaglašenosti	4	3.6
Generator procesne mape sa merama performansi	4	3.6
Analiza uskih grla	3	2.7
<i>PALIA</i> – ugrađene tehnike unapređenja	3	2.7
Analiza paterna	2	1.8
<i>ProM</i> dodatak za predikciju	2	1.8
Interaktivni heuristički algoritam - ugrađene tehnike unapređenja	2	1.8
<i>AITIA-PM</i> – ugrađene tehnike unapređenja	1	0.9

Algoritam za otkrivanje pravila	1	0.9
Algoritam za otkrivanje tranzicionih sistema	1	0.9
Algoritam za procenu vremena čekanja	1	0.9
Analiza distribucije vremena i prostora	1	0.9
Analiza performansi zasnovana na poravnanju	1	0.9
Analiza varijanti toka procesa	1	0.9
Algoritam za otkrivanje odluka	1	0.9
Graf direktnog praćenja	1	0.9
<i>ProM</i> dodatak – Više-perspektivni istraživač procesa	1	0.9
Algoritam za distribuirano i deklarativno otkrivanje procesa	1	0.9
Vizualni induktivni algoritam – prikaz devijacija procesa	1	0.9
Interaktivno otkrivanje poslovnih procesa – ugrađene tehnike unapređenja	1	0.9
Izveštaj o devijacijama	1	0.9
<i>Minit</i> – ugrađene tehnike unapređenja	1	0.9
Objektno-orijentisana analiza performansi	1	0.9
Otkrivanje sekvencijalnih paterna	1	0.9
Perspektiva performansi	1	0.9
Ukupno	110	100.0

Tabela 29 predstavlja učestalost i udeo softverskih alata za unapređenje poslovnih procesa u posmatranim primarnim studijama. Broj primarnih studija koji se posmatra odgovara broju primarnih studija koje su izvršile tip unapređenje poslovnih procesa. Najčešće primenjivani softverski alat za primenu tehnika unapređenja poslovnih procesa je *ProM*, sa udelom od 40%. Zatim sledi softverski alat *Disco* sa udelom od 31.8%, *Celonis* sa udelom od 7.3%, *BupaR* sa udelom od 3.6%, alati razvijeni od strane autora primarnih studija sa udelom od 2.7%, *PALIA* sa udelom od 2.7% i pajton biblioteka *PM4Py* sa udelom od 1.8%. Preostali softverski alati za primenu tehnika unapređenja poslovnih procesa su korišćeni u pojedinačnim slučajevima.

Tabela 29. Učestalost i udeo softverskih alata za unapređenje poslovnih procesa

Softverski alat za unapređenje poslovnih procesa	Učestalost	Udeo u %
<i>ProM</i>	44	40.0
<i>Disco</i>	35	31.8
<i>Celonis</i>	8	7.3
<i>BupaR</i>	4	3.6
Alat razvijen od strane autora primarne studije	3	2.7
<i>PALIA</i>	3	2.7
<i>PM4Py</i>	2	1.8
<i>AITIA-PM</i>	1	0.9
<i>Apromore</i>	1	0.9
<i>EverFlow</i>	1	0.9
Alat za interaktivno otkrivanje poslovnih procesa	1	0.9
<i>Minit</i>	1	0.9
<i>Networkx2</i>	1	0.9

<i>PAFnow</i>	1	0.9
<i>ProDiGen</i>	1	0.9
<i>Rio Diagnostics</i>	1	0.9
<i>SOWCompact</i>	1	0.9
<i>Upflux</i>	1	0.9
Ukupno	110	100.0

Tabela 30 predstavlja učestalost i udeo tehnika za proveru usaglašenosti posmatranih primarnih studija. Broj posmatranih slučajeva odgovara broju primarnih studija koje su primenile tehnike za proveru usaglašenosti. Najčešće korišćena tehnika za proveru usaglašenosti je *ProM* dodatak za proveru usaglašenosti, koji se zasniva na tehnikama poravnavanja modela procesa sa dnevnikom izvršenja događaja. Zatim, u 14.8% posmatranih slučajeva, primenjene su tehnike provere usaglašenosti zasnovane na poravnavanju sadržane u različitim softverskim alatima. Takođe, prilikom primene induktivnog algoritma moguće je odabrati opciju za proveru usaglašenosti, što su autori uradili u 7.4% posmatranih primarnih studija. U nekim slučajevima, autori su manuelno proveravali usaglašenost otkrivenog modela procesa sa unapred utvrđenim pravilima izvršavanja procesa. Preostale tehnike za proveru usaglašenosti su primenjene u pojedinačnim slučajevima i to su: *ProM* dodatak za deklarativnu analizu, *ProM* dodatak – više-perspektivni istraživač procesa, interaktivno otkrivanje poslovnih procesa – ugrađene tehnike unapređenja, opcija za proveru pravila i otkrivanje pseudo-toka procesa.

Tabela 30. Učestalost i udeo tehnika za proveru usaglašenosti

Tehnike za proveru usaglašenosti	Učestalost	Udeo u %
<i>ProM</i> dodatak za proveru usaglašenosti	15	55.6
Opcija za proveru usaglašenosti	4	14.8
Induktivni algoritam – ugrađene tehnike unapređenja	2	7.4
Manualna provera usaglašenosti zasnovana na pravilima	2	7.4
<i>ProM</i> dodatak za deklarativnu analizu	1	3.7
<i>ProM</i> dodatak – Više-perspektivni istraživač procesa	1	3.7
Interaktivno otkrivanje poslovnih procesa – ugrađene tehnike unapređenja	1	3.7
Opcija za proveru pravila	1	3.7
Otkrivanje pseudo-toka procesa	1	3.7
Ukupno	27	100.0

Tabela 31 predstavlja učestalost i udeo softverskih alata za proveru usaglašenosti u posmatranim primarnim studijama. Alat koji je primenjen u 70.4% posmatranih primarnih studija je *ProM*. Preostali softverski alati su primenjeni u mnogo manjoj meri, i to *Celonis* sa udelom od 14.8%, *Disco* sa udelom od 3.7%, manualna primena tehnika sa udelom od 3.7%, pajton biblioteka *PM4Py* sa udelom od 3.7%, *pMineR* sa udelom od 3.7% i alat za interaktivno otkrivanje poslovnih procesa sa udelom od 3.7%.

Tabela 31. Učestalost i udeo softverskih alata za proveru usaglašenosti

Softverski alati za proveru usaglašenosti	Učestalost	Udeo u %
<i>ProM</i>	19	70.4
<i>Celonis</i>	4	14.8
<i>Disco</i>	1	3.7
Manualna primena tehnika	1	3.7
<i>PM4Py</i>	1	3.7

<i>pMineR</i>	1	3.7
Alat za interaktivno otkrivanje poslovnih procesa	1	3.7
Ukupno	27	100.0

Prilikom analize društvenih mreža autori su koristili mogućnosti različitih softverskih alata kako bi otkrili model društvene mreže između resursa koji učestvuju u izvršavanju aktivnosti nekog procesa. Broj slučajeva koji se posmatra odgovara broju primarnih studija unutar kojih je izvršena analiza društvenih mreža. U 53.8% slučajeva, softverski alat unutar kojeg je primenjena tehnika otkrivanja društvenih mreža je *ProM*. Mogućnosti koje pruža pajton biblioteka *PM4Py* korišćene su u 15.4% slučajeva, dok je softverski alat *Minit* korišćen u 7.7% slučajeva.

Tabela 32. Učestalost i udeo softverskih alata za analizu društvenih mreža

Softverski alati za analizu društvenih mreža	Učestalost	Udeo u %
<i>ProM</i>	7	53.8
<i>PM4Py</i>	2	15.4
<i>Minit</i>	1	7.7
Ukupno	13	100.0

4.1.1.4. Pridružene tehnologije

Prilikom procesa ekstrakcije podataka, posmatrane su i tehnologije ili tehnike koje ne pripadaju oblasti automatskog otkrivanja poslovnih procesa, ali su korišćene kako bi se upotpunila i omogućila primena automatskog otkrivanja poslovnih procesa.

Tabela 33 predstavlja tehnologije koje su primenjene u 54 različite primarne studije. Klasterovanje je, kao tehnika istraživanja i eksploatacije podataka, primenjeno u 11 primarnih studija i pokazalo se kao dominantna pridružena tehnika automatskog otkrivanja poslovnih procesa. Sama oblast istraživanja i eksploatacije podataka je navedena u još 2 primarne studije. Deskriptivna statistika je takođe primenjena u velikom broju posmatranih slučajeva, sa učešćem od 14.8%. Mašinsko učenje je kao oblast navedena u 7 primarnih studija, ali su primenjivane i konkretne tehnike i algoritmi, poput NLP u 3 primarne studije, duboke neuronske mreže u 1 primarnoj studiji, otkrivanje podataka iz teksta u 1 primarnoj studiji i regresioni model nasumične šume u 1 primarnoj studiji. Diskretna simulacija događaja je primenjena u 2 primarne studije, kao i simulacija. TDABC (engl. *Time-drive activity-based costing*) je pristup računanja troškova kretanja pacijenta kroz zdravstveni sistem i primenjen je u 2 primarne studije. Grafička evaluacija i pregled (engl. *Graphical Evaluation and Review Technique* – GERT) je pristup projektnog menadžmenta koji se koristi za planiranje kompleksnih projekata i primenjen je u 1 primarnoj studiji. Preostale pridružene tehnike primenjene u pojedinačnim slučajevima su analiza obuhvata podataka, analiza preživljavanja, analiza vrednosti toka podataka, blokčej, digitalni blizanac, eksploratorna faktorska analiza, Globalni pozicioni sistem – GPS, korelacija, *Lean Six Sigma* pristup, robotska automatizacija procesa, statistička kontrola procesa, temporalna digitalna fenotipizacija zasnovana na abstrakciji i veštačka inteligencija.

Tabela 33. Učestalost i udeo pridruženih tehnologija automatskog otkrivanja poslovnih procesa

Pridružene tehnologije	Učestalost	Udeo u %
Klasterovanje	11	20.4
Deskriptivna statistika	8	14.8
Mašinsko učenje - uopšteno	7	13.0
NLP	3	5.6
Diskretna simulacija događaja	2	3.7
Istraživanje i eksploatacija podataka	2	3.7
Simulacija	2	3.7

TDABC pristup	2	3.7
Analiza obuhvata podataka	1	1.9
Analiza preživljavanja	1	1.9
Analiza vrednosti toka podataka	1	1.9
Blokčejn	1	1.9
Digitalni blizanac	1	1.9
Duboke neuronske mreže	1	1.9
Eksploratorna faktorska analiza	1	1.9
GERT pristup	1	1.9
GPS	1	1.9
Korelacija	1	1.9
<i>Lean Six Sigma</i> pristup	1	1.9
Otkrivanje podataka iz teksta	1	1.9
Regresioni model nasumične šume	1	1.9
Robotska automatizacija procesa	1	1.9
Statistička kontrola procesa	1	1.9
Temporalna digitalna fenotipizacija zasnovana na abstrakciji	1	1.9
Veštačka inteligencija	1	1.9
Ukupno	54	100.0

4.1.2. Rezultati ispitivanja hipoteza istraživačkog modela I

Istraživački model 1 sadrži 5 hipoteza, čiji su rezultati ispitivanja predstavljeni u narednim potpoglavljima.

4.1.2.1. Rezultat ispitivanja hipoteze H1: Povezanost vrste industrije sa vrstom poslovnog procesa

Tabela 34 pokazuje rezultate Hi-kvadrat testa i Fišerovog egzaktnog testa primenjenih radi ispitivanja postojanja veze između vrste industrije i vrste poslovnog procesa primarnih studija. Rezultat Hi-kvadrat testa se ne može uvažiti, jer nije ispunjena pretpostavka testa da najviše 20% posmatranih ćelija sme da ima očekivanu frekvenciju manju od 5. Prema tabeli 34 može se zaključiti da 99.2% posmatranih ćelija ima očekivanu učestalost manju od 5.

Stoga se u obzir uzima rezultat Fišerovog egzaktnog testa, koji pokazuje da postoji statistički značajna veza između vrste industrije i vrste poslovnog procesa ($p = 0.000$).

Tabela 34. Rezultat ispitivanja povezanosti vrsta industrije ↔ vrsta poslovnog procesa

Hi-kvadrat test	Vrednost	df	Značajnost (p-vrednost)	Egzaktna značajnost (p-vrednost)
Pirsonov Hi-kvadrat	877.271 ^a	328	0.000	0.000
Racio verovatnoće	421.956	328	0.000	0.000
Fišerov egzaktni test	0.000			0.000
Broj validnih slučajeva	144			

a. 143 ćelije (99.2%) ima očekivani broj manji od 5. Minimalni očekivani broj je 0.01.

Iz prethodno navedenog možemo zaključiti da se **hipoteza H1 prihvata**.

4.1.2.1. Rezultat ispitivanja hipoteze H2: Povezanost vrste industrije sa ciljem analize automatskog otkrivanja poslovnih procesa

Tabela 35 pokazuje rezultate Hi-kvadrat testa i Fišerovog egzaktnog testa primenjenih radi ispitivanja postojanja veze između vrste industrije i sa ciljem istraživanja primarnih studija. Rezultat Hi-kvadrat testa se ne može uvažiti, jer nije ispunjena pretpostavka testa da najviše 20% posmatranih ćelija sme da ima očekivanu frekvenciju manju od 5. Prema tabeli 35 može se zaključiti da 94.9% posmatranih ćelija ima očekivanu učestalost manju od 5.

Stoga se u obzir uzima rezultat Fišerovog egzaktnog testa, koji pokazuje da postoji statistički značajna veza između vrste industrije i cilja istraživanja ($p = 0.000$).

Tabela 35. Rezultat ispitivanja povezanosti vrsta industrije ↔ cilj analize

Hi-kvadrat test	Vrednost	df	Značajnost (p-vrednost)	Egzaktna značajnost (p-vrednost)
Pirsonov Hi-kvadrat	95.815 ^a	80	0.110	.000
Racio verovatnoće	76.149	80	0.601	.000
Fišerov egzaktni test	.000			.000
Broj validnih slučajeva	144			

a. 94 ćelije (94.9%) ima očekivani broj manji od 5. Minimalni očekivani broj je 0.01.

Iz prethodno navedenog možemo zaključiti da se **hipoteza H2 prihvata**.

4.1.2.2. Rezultat ispitivanja hipoteze H3: Povezanost cilja analize automatskog otkrivanja poslovnih procesa sa vrstom poslovnog procesa

Tabela 36 pokazuje rezultate Hi-kvadrat testa i Fišerovog egzaktnog testa primenjenih radi ispitivanja postojanja veze između cilja analize i vrste poslovnog procesa. Rezultat Hi-kvadrat testa se ne može uvažiti, jer nije ispunjena pretpostavka testa da najviše 20% posmatranih ćelija sme da ima očekivanu frekvenciju manju od 5. Prema tabeli 36 može se zaključiti da 99.0% posmatranih ćelija ima očekivanu učestalost manju od 5.

Stoga se u obzir uzima rezultat Fišerovog egzaktnog testa, koji pokazuje da ne postoji statistički značajna veza između vrste industrije i cilja istraživanja ($p = 0.055$).

Tabela 36. Rezultat ispitivanja povezanosti cilj analize ↔ vrsta poslovnog procesa

Hi-kvadrat test	Vrednost	df	Značajnost (p-vrednost)	Egzaktna značajnost (p-vrednost)
Pirsonov Hi-kvadrat	340.615 ^a	369	.853	0.055
Racio verovatnoće	168.983	369	1.000	0.055
Fišerov egzaktni test	.055			0.055
Broj validnih slučajeva	144			

a. 416 ćelija (99.0%) ima očekivani broj manji od 5. Minimalni očekivani broj je 0.01.

Iz prethodno navedenog možemo zaključiti da se **hipoteza H3 odbacuje**.

4.1.2.3. Rezultat ispitivanja hipoteze H4: Povezanost cilja analize automatskog otkrivanja poslovnih procesa sa tipom

Tabela 37 pokazuje rezultate Hi-kvadrat testa primenjenog radi ispitivanja postojanja veze između cilja analize i tipa automatskog otkrivanja poslovnih procesa. Rezultat Hi-kvadrat testa se može uvažiti, jer je ispunjena pretpostavka testa da najviše 20% posmatranih ćelija sme da ima očekivanu frekvenciju manju od 5. Prema tabeli 37 može se zaključiti da 12.5% posmatranih ćelija ima očekivanu učestalost manju od 5.

Rezultat Hi-kvadrat testa pokazuje da postoji statistički značajna veza između cilja analize i tipova automatskog otkrivanja poslovnih procesa ($p = 0.000$).

Tabela 37. Rezultat ispitivanja povezanosti cilj analize ↔ tip automatskog otkrivanja poslovnih procesa

Hi-kvadrat test	Vrednost	df	Značajnost (p-vrednost)
Pirsonov Hi-kvadrat	34.303 ^a	10	0.000
Racio verovatnoće	15.602	10	0.012
Broj validnih slučajeva	144		

a. 18 ćelija (12.5%) ima očekivani broj manji od 5. Minimalni očekivani broj je 0.55.

Iz prethodno navedenog možemo zaključiti da se **hipoteza H4 prihvata**.

4.1.2.4. Rezultat ispitivanja hipoteze H5: Povezanost cilja analize automatskog otkrivanja poslovnih procesa sa primenjenim algoritmima i tehnikama

Kako bi se ispitala hipoteza H5 potrebno je zasebno ispitati povezanost između cilja analize sa algoritmima i tehnikama svakog tipa automatskog otkrivanja poslovnih procesa.

Tabela 38 predstavlja rezultate Hi-kvadrat testa i Fišerovog egzaktnog testa primenjenih radi ispitivanja postojanja veze između cilja analize i algoritama tipa otkrivanje poslovnih procesa. Rezultat Hi-kvadrat testa se ne može uvažiti, jer nije ispunjena pretpostavka testa da najviše 20% posmatranih ćelija sme da ima očekivanu frekvenciju manju od 5. Prema tabeli 38 može se zaključiti da 98.8% posmatranih ćelija ima očekivanu učestalost manju od 5.

Stoga se u obzir uzima rezultat Fišerovog egzaktnog testa, koji pokazuje da postoji statistički značajna veza između cilja analize i algoritama za otkrivanje poslovnih procesa ($p = 0.000$).

Tabela 38. Rezultat ispitivanja povezanosti cilj analize ↔ algoritmi za otkrivanje poslovnih procesa

Hi-kvadrat test	Vrednost	df	Značajnost (p-vrednost)	Egzaktna značajnost (p-vrednost)
Pirsonov Hi-kvadrat	344.823 ^a	300	.038	0.000
Racio verovatnoće	120.900	300	1.000	0.000
Fišerov egzaktni test	0.000			0.000
Broj validnih slučajeva	144			

a. 337 ćelija (98.8%) ima očekivani broj manji od 5. Minimalni očekivani broj je 0.01.

Tabela 39 predstavlja rezultate Hi-kvadrat testa i Fišerovog egzaktnog testa primenjenih radi ispitivanja postojanja veze između cilja analize i tehnika unapređenja poslovnih procesa. Rezultat Hi-kvadrat testa se ne može uvažiti, jer nije ispunjena pretpostavka testa da najviše 20% posmatranih ćelija sme da ima očekivanu frekvenciju manju od 5. Prema tabeli 38 može se zaključiti da 99.0% posmatranih ćelija ima očekivanu učestalost manju od 5.

Stoga se u obzir uzima rezultat Fišerovog egzaktnog testa, koji pokazuje da postoji statistički značajna veza između cilja analize i tehnika unapređenja poslovnih procesa ($p = 0.000$).

Tabela 39. Rezultat ispitivanja povezanosti cilj analize ↔ tehnike unapređenja poslovnih procesa

Hi-kvadrat test	Vrednost	df	Značajnost (p-vrednost)	Egzaktna značajnost (p-vrednost)
Pirsonov Hi-kvadrat	455.158 ^a	360	.000	0.000
Racio verovatnoće	144.782	360	1.000	0.000
Fišerov egzaktni test	.000			0.000
Broj validnih slučajeva	144			

a. 403 ćelije (99.0%) ima očekivani broj manji od 5. Minimalni očekivani broj je 0.01.

Tabela 40 predstavlja rezultate Hi-kvadrat testa i Fišerovog egzaktnog testa primenjenih radi ispitivanja postojanja veze između cilja analize i tehnika provere usaglašenosti poslovnih procesa. Rezultat Hi-kvadrat testa se ne može uvažiti, jer nije ispunjena pretpostavka testa da najviše 20% posmatranih ćelija sme da ima očekivanu frekvenciju manju od 5. Prema tabeli 40 može se zaključiti da 99.0% posmatranih ćelija ima očekivanu učestalost manju od 5.

Stoga se u obzir uzima rezultat Fišerovog egzaktnog testa, koji pokazuje da postoji statistički značajna veza između cilja analize i tehnika provere usaglašenosti procesa ($p = 0.000$).

Tabela 40. Rezultat ispitivanja povezanosti cilj analize ↔ tehnike provere usaglašenosti procesa

Hi-kvadrat test	Vrednost	df	Značajnost (p-vrednost)	Egzaktna značajnost (p-vrednost)
Pirsonov Hi-kvadrat	455.158a	360	.000	0.000
Racio verovatnoće	144.782	360	1.000	0.000
Fišerov egzaktni test	.000			0.000
Broj validnih slučajeva	144			

a. 403 ćelije (99.0%) ima očekivani broj manji od 5. Minimalni očekivani broj je 0.01.

Tabela 41 predstavlja rezultate Hi-kvadrat testa i Fišerovog egzaktnog testa primenjenih radi ispitivanja postojanja veze između cilja analize i tehnika analize društvenih mreža. Rezultat Hi-kvadrat testa se ne može uvažiti, jer nije ispunjena pretpostavka testa da najviše 20% posmatranih ćelija sme da ima očekivanu frekvenciju manju od 5. Prema tabeli 41 može se zaključiti da 99.0% posmatranih ćelija ima očekivanu učestalost manju od 5.

Stoga se u obzir uzima rezultat Fišerovog egzaktnog testa, koji pokazuje da postoji statistički značajna veza između cilja analize i tehnika analize društvenih mreža ($p = 0.000$).

Tabela 41. Rezultat ispitivanja povezanosti cilj analize ↔ tehnike analize društvenih mreža

Hi-kvadrat test	Vrednost	df	Značajnost (p-vrednost)	Egzaktna značajnost (p-vrednost)
Pirsonov Hi-kvadrat	455.158a	360	.000	0.000
Racio verovatnoće	144.782	360	1.000	0.000
Fišerov egzaktni test	.000			0.000
Broj validnih slučajeva	144			

a. 403 ćelije (99.0%) ima očekivani broj manji od 5. Minimalni očekivani broj je 0.01.

Tabela 42 predstavlja rezultate Hi-kvadrat testa i Fišerovog egzaktnog testa primenjenih radi ispitivanja postojanja veze između cilja analize i tehnika predviđanja. Rezultat Hi-kvadrat testa se ne može uvažiti, jer nije ispunjena pretpostavka testa da najviše 20% posmatranih ćelija sme da ima očekivanu frekvenciju manju od 5. Prema tabeli 41 može se zaključiti da 88.9% posmatranih ćelija ima očekivanu učestalost manju od 5.

Stoga se u obzir uzima rezultat Fišerovog egzaktnog testa, koji pokazuje da postoji statistički značajna veza između cilja analize i tehnika predviđanja ($p = 0.000$).

Tabela 42. Rezultat ispitivanja povezanosti cilj analize ↔ tehnike predviđanja

Hi-kvadrat test	Vrednost	df	Značajnost (p-vrednost)	Egzaktna značajnost (p-vrednost)
Pirsonov Hi-kvadrat	325.434a	55	0.000	0.000
Racio verovatnoće	314.427	55	0.000	0.000
Fišerov egzaktni test	0.000			0.000
Broj validnih slučajeva	246			

a. 64 ćelije (88.9%) ima očekivani broj manji od 5. Minimalni očekivani broj je 0.01.

Iz prethodno navedenog možemo zaključiti da se **hipoteza H5 prihvata**.

4.1.3. Tabela kontingencije odnosa između vrste proizvodnje, cilja analize i tipa automatskog otkrivanja poslovnih procesa

Na osnovu potvrđenih hipoteza o povezanosti između vrste industrije i cilja analize, kao i cilja analize i tipova automatskog otkrivanja poslovnih procesa, unakrsnom analizom je kreirana tabela kontingencije i predstavljena u tabeli 43. Tabela kontingencije odnosa između vrste proizvodnje, cilja analize i tipa automatskog otkrivanja poslovnih procesa unutar svake ćelije prikazuje učestalost primene određenog tipa automatskog otkrivanja poslovnih procesa, koji je određen na osnovu industrije u kojoj se analiza sprovodila i cilja analize automatskog otkrivanja poslovnih procesa.

Tabela 43. Tabela kontingencije odnosa između vrste proizvodnje, cilja analize i tipa automatskog otkrivanja poslovnih procesa

Industrija	Cilj analize	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	Analiza društvenih mreža	Prediktivno AOPP
Proizvodnja	Otkrivanje i unapređenje poslovnog procesa	13	13	3	2	1
	Pronalazak anomalija i nepredviđenih aktivnosti	3	3	1	0	0
	Analiza vremena čekanja	3	3	0	0	0
	Predviđanje ponašanja procesa	3	2	0	1	2
	Automatizacija poslovnog procesa	1	1	0	0	0
	Otkrivanje poslovnih pravila	0	1	0	0	0
	Analiza promena procesa	1	0	0	0	0
Zdravstvo	Otkrivanje i unapređenje poslovnog procesa	34	29	6	3	0
	Poređenje poslovnih procesa	4	3	2	0	0
	Predviđanje ponašanja procesa	3	1	0	0	1
	Analiza resursa	2	1	0	1	0
	Analiza vremena čekanja	2	2	0	0	0
	Analiza promena procesa	1	1	0	0	0
Obrazovanje	Otkrivanje i unapređenje poslovnog procesa	17	13	4	0	0
	Otkrivanje uzroka problema	1	1	0	0	0
	Predviđanje ponašanja procesa	1	0	1	0	0
	Pronalazak anomalija i nepredviđenih aktivnosti	1	1	0	0	0

Računovodstvo, kontroling, bankarstvo i osiguranje	Automatizacija poslovnog procesa	1	0	0	0	0
	Poređenje poslovnih procesa	1	1	0	0	0
	Otkrivanje i unapređenje poslovnog procesa	3	2	0	1	0
	Predviđanje ponašanja procesa	1	0	0	0	1
	Pronalazak anomalija i nepredviđenih aktivnosti	1	1	0	0	0
	Automatizacija poslovnog procesa	1	0	0	0	0
	Poređenje poslovnih procesa	1	1	1	0	0
Informaciono- komunikacione tehnologije	Otkrivanje i unapređenje poslovnog procesa	13	8	1	1	0
	Pronalazak anomalija i nepredviđenih aktivnosti	6	4	3	0	0
	Predviđanje ponašanja procesa	1	0	0	0	0
Logistika	Otkrivanje i unapređenje poslovnog procesa	3	3	0	0	0
	Analiza promena procesa	1	0	1	0	0
	Predviđanje ponašanja procesa	0	0	0	0	1
	Analiza resursa	1	1	1	0	0
Javni servisi i usluge	Otkrivanje i unapređenje poslovnog procesa	6	6	1	0	0
	Otkrivanje uzroka problema	0	1	0	0	0
	Pronalazak anomalija i nepredviđenih aktivnosti	1	1	0	1	0
Građevina	Otkrivanje i unapređenje poslovnog procesa	3	3	1	2	0
	Predviđanje ponašanja procesa	2	2	0	0	2
Energija	Otkrivanje i unapređenje poslovnog procesa	1	1	1	1	0

4.2. Rezultati istraživačkog modela II

Istraživački model II je imao za cilj da ispita hipotezu koja se tiče postojanja mogućnosti određivanja pogodne tehnike čišćenja dnevnika izvršenja događaja u zavisnosti od problema kvaliteta

podataka. Merni instrument korišćen za prikupljanje podataka je upitnik. Naredna potpoglavlja predstavljaju rezultate primenjene metodologije obrade podataka nad prikupljenim podacima, kao što je navedeno u potpoglavlju 3.2.3.

4.2.1. Percipirani značaj i učestalost susretanja problema kvaliteta dnevnika izvršenja događaja

Rezultati primene deskriptivne statistike nad percipiranim značajem problema kvaliteta dnevnika izvršenja događaja, merenim putem Likertove skale, predstavljeni su procentualno u tabeli 44.

Tabela 44. Percipirani značaj problema kvaliteta podataka dnevnika izvršenja događaja

Problemi kvaliteta podataka	Percipirani značaj u %				
	Nije značajan	Malo značajan	Umereno značajan	Značajan	Veoma značajan
Nedostajući podaci: Slučaj	4.0	15.8	21.8	28.7	29.7
Nedostajući podaci: Događaj (Raštrkani događaj)	0	5.0	20.3	47.5	27.2
Nedostajući podaci: Veza (Neuhvatljiv slučaj)	1.0	6.9	15.3	36.6	40.1
Nedostajući podaci: Naziv aktivnosti	5.4	10.9	31.7	22.8	29.2
Nedostajući podaci: Atribut slučaja/događaja	0	15.8	45.0	25.7	13.4
Nedostajući podaci: Vremenska odrednica	0.5	3.0	4.5	18.8	73.3
Nedostajući podaci: Resurs	6.9	25.7	30.2	28.7	8.4
Netačni podaci: Slučaj	1.0	15.8	18.3	41.1	23.8
Netačni podaci: Događaj	0	5.0	19.8	44.1	31.2
Netačni podaci: Veza (Raštrkani slučaj)	1.0	8.4	21.8	37.6	31.2
Netačni podaci: Naziv aktivnosti (Zagađeni/Iskrivljeni naziv atributa)	2.0	19.3	35.1	22.3	21.3
Netačni podaci: Atribut slučaja/događaja	1.0	11.4	31.7	41.6	14.4
Netačni podaci: Vremenska odrednica (Hvatanje događaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj)	0	8.9	11.9	27.7	51.5
Netačni podaci: Resurs (Zagađeni naziv atributa)	5.9	19.8	38.1	28.2	7.9
Neprecizni podaci: Veza	4.0	13.4	29.7	37.6	15.3
Neprecizni podaci: Naziv aktivnosti (Istoimeni naziv atributa)	3.0	11.9	40.1	31.2	13.9
Neprecizni podaci: Atribut slučaja/događaja (Sinonimni naziv atributa)	2.0	18.3	33.2	37.1	9.4
Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj)	1.0	5.0	18.3	31.2	44.6
Neprecizni podaci: Resurs	7.9	29.2	34.2	22.8	5.9
Irelevantni podaci: Slučaj	15.3	25.2	29.7	21.8	7.9

Irelevantni podaci: Događaj (Hvatanje događaja zasnovano na obrascu, Kolateralni događaj)	16.3	30.2	28.2	17.8	7.4
Količina podataka, nivo detaljnosti događaja, kompleksnost	1.0	6.9	23.8	39.1	29.2

Rezultati primene deskriptivne statistike nad percipiranom učestalošću susretanja problema kvaliteta dnevnika izvršenja događaja, merenim putem Likertove skale, predstavljeni su procentualno u tabeli 45.

Tabela 45. Percipirana učestalost problema kvaliteta podataka dnevnika izvršenja događaja

Problemi kvaliteta podataka	Percipirana učestalost u %				
	Nikad	Retko	Povremeno	Često	Veoma često
Nedostajući podaci: Slučaj	5.0	31.7	34.2	19.8	9.4
Nedostajući podaci: Događaj (Raštrkani događaj)	1.0	22.3	29.7	38.6	8.4
Nedostajući podaci: Veza (Neuhvatljiv slučaj)	7.4	26.7	36.1	24.3	5.4
Nedostajući podaci: Naziv aktivnosti	7.4	35.6	29.2	21.8	5.9
Nedostajući podaci: Atribut slučaja/događaja	3.0	17.8	41.6	29.7	7.9
Nedostajući podaci: Vremenska odrednica	11.9	38.1	26.2	13.4	10.4
Nedostajući podaci: Resurs	3.0	23.8	33.7	27.7	11.9
Netačni podaci: Slučaj	11.9	40.6	25.7	17.8	4.0
Netačni podaci: Događaj	4.5	38.6	32.7	21.3	3.0
Netačni podaci: Veza (Raštrkani slučaj)	6.4	40.6	34.7	12.4	5.9
Netačni podaci: Naziv aktivnosti (Zagađeni/Iskrivljeni naziv atributa)	5.9	33.2	37.6	17.8	5.4
Netačni podaci: Atribut slučaja/događaja	3.5	33.2	42.1	16.8	4.5
Netačni podaci: Vremenska odrednica (Hvatanje događaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj)	6.9	31.7	21.3	27.2	12.9
Netačni podaci: Resurs (Zagađeni naziv atributa)	8.9	40.6	34.2	13.4	3.0
Neprecizni podaci: Veza	5.9	39.6	35.6	17.8	1.0
Neprecizni podaci: Naziv aktivnosti (Istoimeni naziv atributa)	5.9	29.2	38.6	21.8	4.5
Neprecizni podaci: Atribut slučaja/događaja (Sinonimni naziv atributa)	5.4	31.2	37.1	19.8	6.4
Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj)	3.5	32.2	35.1	15.3	13.9
Neprecizni podaci: Resurs	10.4	34.2	34.7	16.8	4.0
Irelevantni podaci: Slučaj	5.9	36.6	25.7	22.8	8.9

Irelevantni podaci: Događaj (Hvatanje događaja zasnovano na obrascu, Kolateralni događaj)	5.0	32.2	32.2	22.3	8.4
Količina podataka, nivo detaljnosti događaja, kompleksnost	5.4	13.9	29.2	37.1	14.4

4.2.2. Percipirani značaj i učestalost primene tehnika čišćenja dnevnika izvršenja događaja

Rezultati primene deskriptivne statistike nad percipiranim značajem primene tehnika čišćenja dnevnika izvršenja događaja, merenim putem Likertove skale, prikazani su procentualno u tabeli 46.

Tabela 46. Percipirani značaj primene tehnika čišćenja dnevnika izvršenja događaja

Tehnike čišćenja dnevnika izvršenja događaja	Percipirani značaj u %				
	Nije značajna	Malo značajna	Umereno značajna	Značajna	Veoma značajna
Klasterovanje putanja	5.4	8.9	35.1	36.6	13.9
Tehnike popravke dnevnika izvršenja događaja	8.4	22.8	24.3	32.7	11.9
Filtriranje putanje/događaja	3	3.5	11.9	39.6	42.1
Abstrakcija događaja	5	2.5	37.6	28.2	26.7
Mašinsko učenje, veštačka inteligencija i istraživanje i eksploatacija podataka	10.4	20.3	32.7	26.2	10.4
Tehnike poravnavanja	10.4	21.8	38.1	22.3	7.4
Ugrađena obrada podataka	8.9	16.3	22.8	33.7	18.3

Rezultati primene deskriptivne statistike nad percipiranom učestalošću primene tehnika čišćenja dnevnika izvršenja događaja, merenim putem Likertove skale, prikazani su procentualno u tabeli 47.

Tabela 47. Percipirana učestalost primene tehnika čišćenja dnevnika izvršenja događaja

Tehnike čišćenja dnevnika izvršenja događaja	Percipirana učestalost u %				
	Nikad	Retko	Povremeno	Često	Veoma često
Klasterovanje putanja	14.9	20.3	33.2	20.8	10.9
Tehnike popravke dnevnika izvršenja događaja	20.3	22.8	31.2	21.8	4
Filtriranje putanje/događaja	4.5	6.4	17.3	32.7	39.1
Abstrakcija događaja	8.4	16.3	30.2	24.8	20.3
Mašinsko učenje, veštačka inteligencija i istraživanje i eksploatacija podataka	24.8	20.8	23.8	19.8	10.9
Tehnike poravnavanja	26.7	23.8	23.8	20.3	5.4

Ugrađena obrada podataka	19.8	23.3	18.8	21.8	16.3
--------------------------	------	------	------	------	------

4.2.3. IPA analiza

Naredna potpoglavlja predstavljaju rezultate IPA analize, sprovedene prema prethodno utvrđenoj metodologiji, radi ispitivanja odnosa između percipiranog značaja i učestalosti susretanja odnosno primene posmatranih dimenzija IM2. Prva IPA analiza se odnosi na analizu odnosa percipiranog značaja i učestalosti susretanja problema kvaliteta podataka dnevnika izvršenja događaja, dok se druga IPA analiza odnosi na analizu odnosa percipiranog značaja i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja.

4.2.3.1. IPA analiza percipiranog značaja i učestalosti susretanja problema kvaliteta podataka dnevnika izvršenja događaja

Primenom IPA analize, izračunate su aritmetičke srednje vrednosti percipiranog značaja i učestalosti za svaku stavku dimenzije problema kvaliteta dnevnika izvršenja događaja, merene putem Likertove skale.

Na osnovu tabele 48 može se zaključiti da su najznačajniji problemi kvaliteta oni koji su imali srednju vrednost percipiranog značaja iznad 4.00, a odnose se na nedostajuće, netačne i irelevantne vremenske odrednice dnevnika podataka, zatim na nedostajuće podatke o vezi između slučaja i događaja, kao i netačne podatke o događajima.

Problemi kvaliteta podataka koji su imali srednju vrednost percipiranog značaja između 3.00 i 3.99 su: nedostajući podaci o događajima; netačni podaci o vezi između slučaja i događaja; problem količine podataka, nivo detaljnosti događaja i kompleksnosti; netačni podaci o resursima; netačni i nedostajući podaci o slučaju; nedostajući podaci o nazivu aktivnosti; netačni podaci o atributu slučaja/događaja; neprecizni podaci o vezi između slučaja i događaja; netačni i neprecizni podaci o nazivu aktivnosti; nedostajući podaci o vezi između slučaja i događaja i nedostajući podaci o resursima procesa.

Najmanje značajni problemi kvaliteta podataka sa srednjom vrednošću manjom od 3.00 su problemi koji se tiču nepreciznih podataka o resursima procesa i irelevantnim podacima o slučajevima i događajima.

Ukoliko se posmatra kolona učestalost u tabeli 48 može se zaključiti da su najučestaliji problemi kvaliteta podataka oni koji su imali srednju vrednost između 3.00 i 3.50, a to su: problem količine podataka, nivo detaljnosti događaja i kompleksnosti; nedostajući podaci o događaju; nedostajući podaci o atributu slučaja/događaja; nedostajući podaci o resursima procesa; netačna vremenska odrednica i neprecizna vremenska odrednica.

Preostali problemi kvaliteta podataka navedeni u tabeli 48 imaju srednju vrednost učestalosti manju od 3.00.

Tabela 48. Aritmetička srednja vrednost percipiranog značaja i učestalosti problema kvaliteta dnevnika izvršenja događaja

	Značaj		Učestalost	
Problemi kvaliteta podataka	Sr. V. ¹	Std. D ²	Sr. V. ¹	Std. D ²
Nedostajući podaci: Slučaj	3.64	1.177	2.97	1.046
Nedostajući podaci: Događaj (Raštrkani događaj)	3.97	.822	3.31	.945
Nedostajući podaci: Veza (Neuhvatljiv slučaj)	4.08	.959	2.94	1.013
Nedostajući podaci: Naziv aktivnosti	3.59	1.173	2.83	1.042

Nedostajući podaci: Atribut slučaja/događaja	3.37	.906	3.22	.932
Nedostajući podaci: Vremenska odrednica	4.61	.753	2.72	1.156
Nedostajući podaci: Resurs	3.06	1.077	3.22	1.033
Netačni podaci: Slučaj	3.71	1.031	2.61	1.036
Netačni podaci: Događaj	4.01	.843	2.80	.927
Netačni podaci: Veza (Raštrkani slučaj)	3.90	.974	2.71	.972
Netačni podaci: Naziv aktivnosti (Zagađeni/Iskrivljeni naziv atributa)	3.42	1.086	2.84	.971
Netačni podaci: Atribut slučaja/događaja	3.57	.907	2.86	.895
Netačni podaci: Vremenska odrednica (Hvatanje događaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj)	4.22	.973	3.07	1.176
Netačni podaci: Resurs (Zagađeni naziv atributa)	3.12	1.012	2.61	.931
Neprecizni podaci: Veza	3.47	1.033	2.68	.869
Neprecizni podaci: Naziv aktivnosti (Istoimeni naziv atributa)	3.41	.969	2.90	.959
Neprecizni podaci: Atribut slučaja/događaja (Sinonimni naziv atributa)	3.34	.949	2.91	.991
Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj)	4.13	.950	3.04	1.083
Neprecizni podaci: Resurs	2.90	1.034	2.70	.999
Irelevantni podaci: Slučaj	2.82	1.172	2.92	1.090
Irelevantni podaci: Događaj (Hvatanje događaja zasnovano na obrascu, Kolateralni događaj)	2.70	1.160	2.97	1.041
Količina podataka, nivo detaljnosti događaja, kompleksnost	3.89	.942	3.41	1.067

¹ Aritmetička srednja vrednost

² Standardna devijacija

Tabela 48 predstavlja izračunate srednje vrednosti, koje su potom primenjene za kreiranje dijagrama rasprostranjenosti problema kvaliteta, predstavljenog na slici 19. Dijagram rasprostranjenosti, kroz horizontalnu, x-osu predstavlja aritmetičku srednju vrednost značaja posmatranih stavki, a kroz vertikalnu, y-osu predstavlja aritmetičku srednju vrednost učestalosti pojavljivanja tih stavki.

Prema slici 19, problemi kvaliteta podataka (P) koji su procenjeni kao visoko značajni i učestali, odnosno pripadaju kvadrantu „Visok prioritet” su:

P1: Nedostajući podaci: Slučaj;

P2: Nedostajući podaci: Događaj (Raštrkani događaj);

P3: Nedostajući podaci: Veza (Neuhvatljiv slučaj);

P13: Netačni podaci: Vremenska odrednica (Hvatanje događaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj);

P18: Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj) i

P22: Količina podataka, nivo detaljnosti događaja, kompleksnost.

Zatim, problemi kvaliteta podataka koji su procenjeni kao visoko značajni, ali retki, i pripadaju kvadrantu „Tako nastavi” su sledeći:

P4: Nedostajući podaci: Naziv aktivnosti;

P6: Nedostajući podaci: Vremenska odrednica;

P8: Netačni podaci: Slučaj;

P9: Netačni podaci: Događaj;

P10: Netačni podaci: Veza (Raštrkani slučaj) i

P12: Netačni podaci: Atribut slučaja/događaja.

Problemi kvaliteta podataka koji su niskog značaja, ali se često susreću i pripadaju kvadrantu „Srednji prioritet” su sledeći:

P5: Nedostajući podaci: Atribut slučaja/događaja;

P7: Nedostajući podaci: Resurs;

P17: Neprecizni podaci: Atribut slučaja/događaja (Sinonimni naziv atributa);

P20: Irelevantni podaci: Slučaj i

P21: Irelevantni podaci: Događaj (Hvatanje događaja zasnovano na obrascu, Kolateralni događaj).

Poslednje, problemi kvaliteta podataka koji su niskog značaja, ne susreću se često i pripadaju kvadrantu „Nizak prioritet” su sledeći:

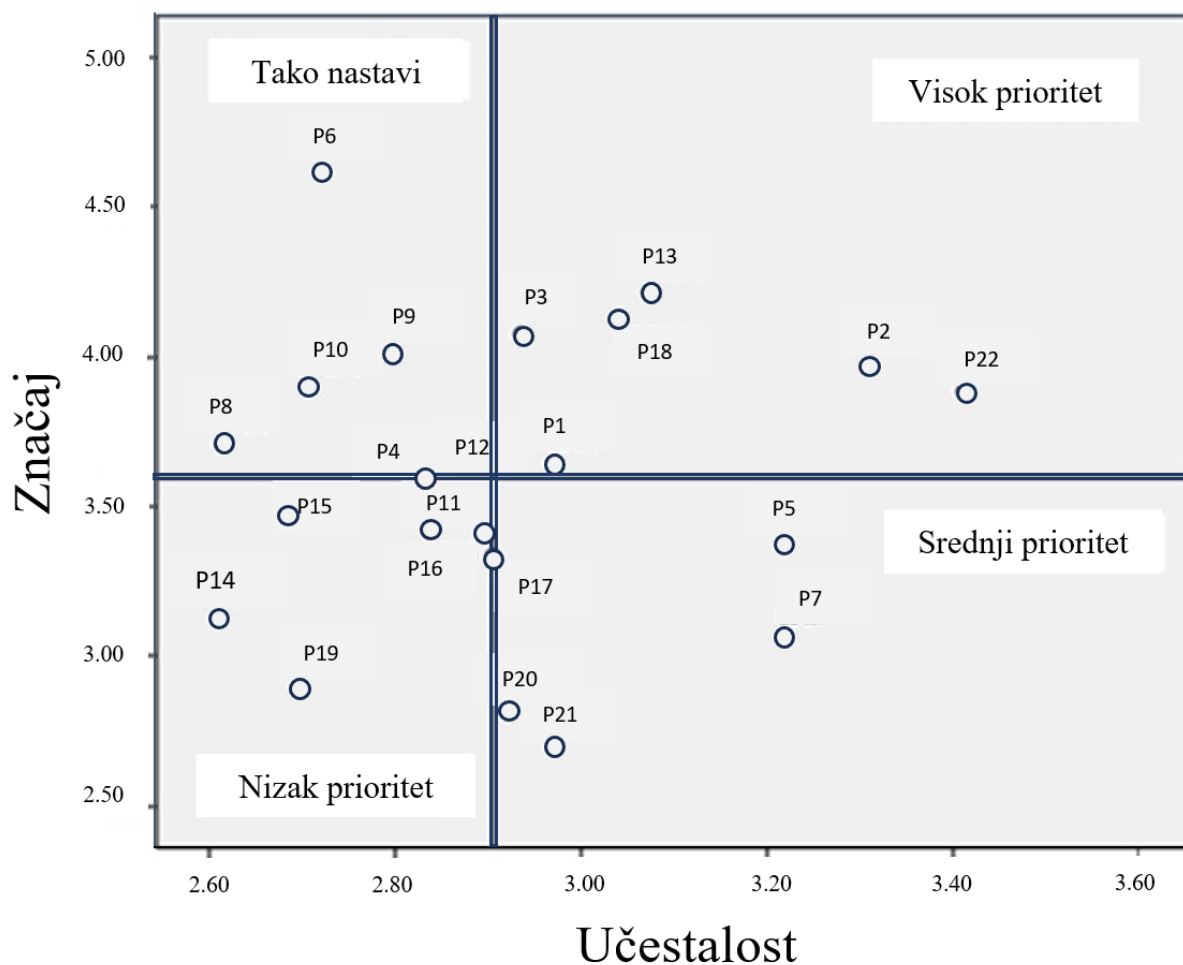
P11: Netačni podaci: Naziv aktivnosti (Zagađeni/Iskrivljeni naziv atributa);

P14: Netačni podaci: Resurs (Zagađeni naziv atributa);

P15: Neprecizni podaci: Veza;

P16: Neprecizni podaci: Naziv aktivnosti (Istoimeni naziv atributa) i

P19: Neprecizni podaci: Resurs.



Slika 19. Dijagram rasprostranjenosti problema kvaliteta podataka dnevnika izvršenja događaja na osnovu percipiranog značaja i učestalosti susretanja

4.2.3.2. IPA analiza percipiranog značaja i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja

Primenom IPA analize, izračunate su aritmetičke srednje vrednosti percipiranog značaja i učestalosti za svaku stavku dimenzije tehnika čišćenja dnevnika izvršenja događaja, merene putem Likertove skale i predstavljene u tabeli 49.

Posmatrajući kolonu značaj u tabeli 49, može se zaključiti da je tehnika filtriranja putanje/događaja posmatrana kao najznačajnija, sa visokom srednjom vrednošću značaja od 4.14. Preostale tehnike poređane prema značaju su sledeće: abstrakcija događaja, sa srednjom vrednošću značaja od 3.69; klasterovanje putanja, sa srednjom vrednošću značaja od 3.45; ugrađena obrada podataka, sa srednjom vrednošću značaja od 3.36; mašinsko učenje, veštačka inteligencija i istraživanje i eksploatacija podataka, sa srednjom vrednošću značaja od 3.06; tehnike popravke dnevnika izvršenja događaja, sa srednjom vrednošću značaja od 3.17 i tehnike poravnavanja, sa srednjom vrednošću značaja od 2.95.

Ukoliko se posmatra kolona učestalosti u tabeli 49, može se zaključiti da percipirani značaj tehnika odgovara njihovoj učestalosti primene, što će se posle pokazati u dijagramu rasprostranjenosti. Tehnika filtriranja putanje/događaja je najčešće primenjena tehnika, sa srednjom vrednošću učestalosti primene od 3.96. Zatim, poređene prema učestalosti primene, slede: abstrakcija događaja, sa srednjom vrednošću učestalosti primene od 3.32; klasterovanje putanja, sa srednjom vrednošću učestalosti primene od 2.93; ugrađena obrada podataka, sa srednjom vrednošću učestalosti primene od 2.92; mašinsko učenje, veštačka inteligencija i istraživanje i eksploatacija podataka, sa srednjom vrednošću

učestalosti primene od 2.71; tehnike popravke dnevnika izvršenja događaja, sa srednjom vrednošću učestalosti primene od 2.66 i tehnike poravnavanja, sa srednjom vrednošću učestalosti primene od 2.54.

Tabela 49. Aritmetička srednja vrednost percipiranog značaja i učestalosti tehnika čišćenja dnevnika izvršenja događaja

Tehnike čišćenja dnevnika izvršenja događaja	Značaj		Učestalost	
	Sr. V. ¹	Std. D ²	Sr. V. ¹	Std. D ²
Klasterovanje putanja	3.45	1.017	2.93	1.201
Tehnike popravke dnevnika izvršenja događaja	3.17	1.160	2.66	1.144
Filtriranje putanje/događaja	4.14	.964	3.96	1.108
Abstrakcija događaja	3.69	1.049	3.32	1.210
Mašinsko učenje, veštačka inteligencija i istraživanje i eksploatacija podataka	3.06	1.140	2.71	1.326
Tehnike poravnavanja	2.95	1.075	2.54	1.234
Ugrađena obrada podataka	3.36	1.211	2.92	1.378

¹ Aritmetička srednja vrednost

² Standardna devijacija

Slika 20 predstavlja dijagram rasprostranjenosti tehnika čišćenja dnevnika izvršenja događaja (T), nastalog na osnovu aritmetičkih srednjih vrednosti njihovog percipiranog značaja i učestalosti primene, predstavljenih u tabeli 49. Prema dijagramu rasprostranjenosti, može se zaključiti da su tehnike čišćenja dnevnika izvršenja događaja raspoređene u dva kvadranta, „Tako nastavi“ i „Nizak prioritet“, sa jednom tehnikom koja se nalazi na preseku svih kvadranta.

Tehnike koje su procenjene kao visoko značajne i učestale i pripadaju kvadrantu „Tako nastavi“ su sledeće:

T1: Klasterovanje putanja;

T3: Filtriranje putanje/događaja i

T4: Abstrakcija događaja.

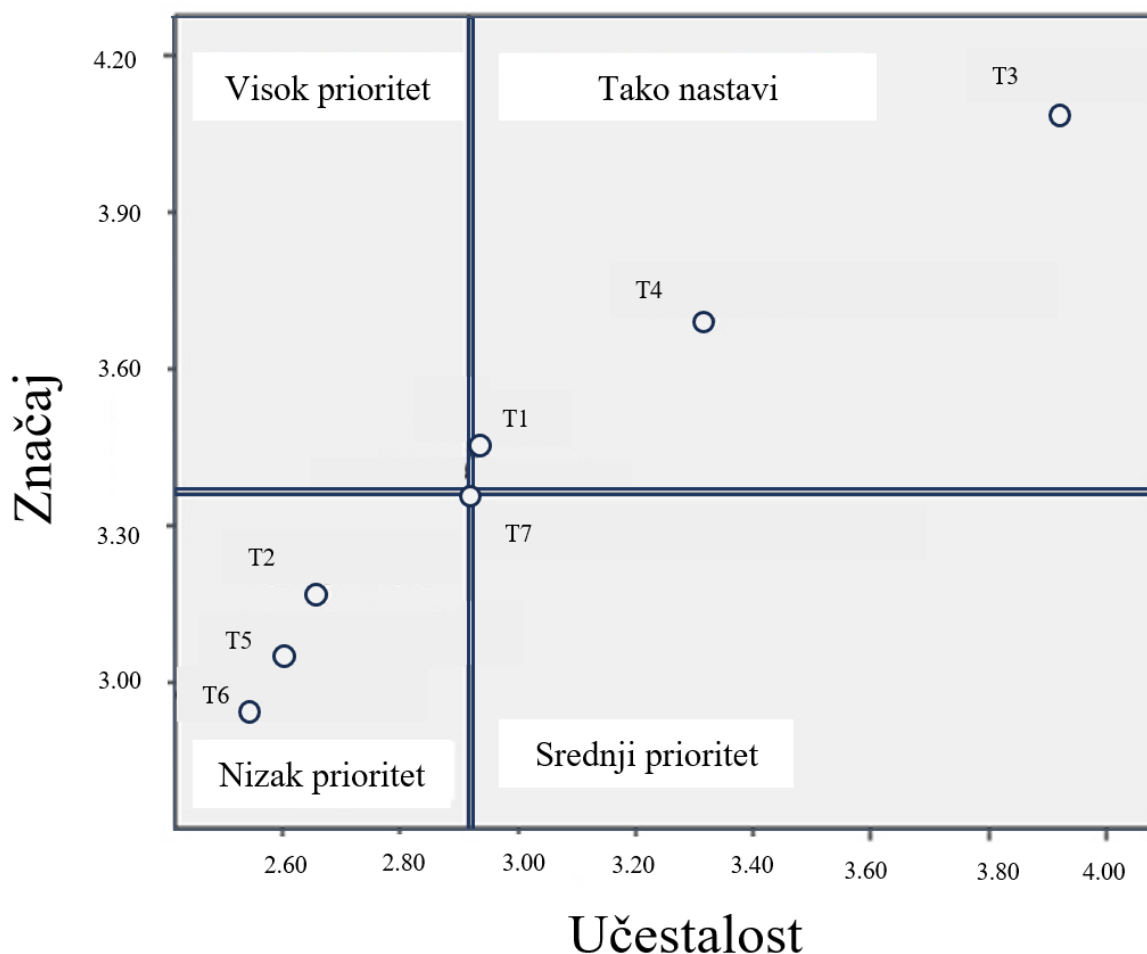
Tehnike koje su procenjene kao niskog značaja i ne primenjuju se često i pripadaju kvadrantu „Nizak prioritet“ su sledeće:

T2: Tehnike popravke dnevnika izvršenja događaja;

T5: Mašinsko učenje, veštačka inteligencija i istraživanje i eksploatacija podataka i

T6: Tehnike poravnavanja.

Tehnike ugrađene obrade podataka (T7) se nalaze na preseku svih kvadranta dijagrama rasprostranjenosti. Kvadrantu „Visok prioritet“ ne pripada nijedna tehnika čišćenja dnevnika izvršenja događaja.



Slika 20. Dijagram rasprostranjenosti tehnika čišćenja podataka dnevnika izvršenja događaja na osnovu percipiranog značaja i učestalosti primene

4.2.4. Analiza varijansi

Kako bi se ispitala razlike između ispitanika u pogledu njihove uloge u oblasti automatskog otkrivanja poslovnih procesa i iskustva koje imaju u oblasti, a u odnosu na njihovu percepciju i učestalost susreta sa posmatranim dimenzijama istraživanja, primenjena je analiza varijansi – Anova. Naredna potpoglavlja predstavljaju dobijene rezultate.

4.2.4.1. Razlike u značaju koji ispitanici različitih uloga pridaju problemima kvaliteta podataka dnevnika izvršenja događaja i tehnikama čišćenja dnevnika izvršenja događaja

Prva analiza varijansi je sprovedena kako bi se ispitala razlika u značaju koji ispitanici različitih uloga pridaju problemima kvaliteta podataka dnevnika izvršenja događaja, gde je rezultat testa pokazao da postoji statistički značajna razlika. Tabela 50. sadrži stavke dimenzije problema kvaliteta podataka dnevnika izvršenja događaja kod kojih postoji statistički značajna razlika u pogledu značaja koji ispitanici pridaju određenom problem, u zavisnosti od njihove uloge. Ispitanici mogu imati tri uloge: istraživač, praktikant ili oba. Tabela 50. sadrži F vrednost testa, p-vrednost koja pokazuje statističku značajnost i post hoc LSD test koji ukazuje na konkretne razlike.

LSD post hoc test je pokazao da praktikanti pridaju manji značaj problemima nedostajućih naziva aktivnosti, netačnih naziva aktivnosti i nepreciznih naziva aktivnosti. Sa druge strane, ispitanici koji su i istraživači i praktikanti pridaju veći značaj problemima nedostajućih podataka o resursima i netačnim podacima o vezi između događaja i slučaja procesa.

Tabela 50. Razlike u značaju koji ispitanici različitih uloga pridaju problemima kvaliteta podataka dnevnika izvršenja događaja

Problemi kvaliteta podataka	F	Značajnost (p-vrednost)	Post hok LSD test*
Nedostajući podaci: Naziv aktivnosti	5.350	.005	2<1,3
Nedostajući podaci: Resurs	3.994	.020	3>2,1
Netačni podaci: Veza (Raštrkani slučaj)	3.820	.024	3>2,1
Netačni podaci: Naziv aktivnosti (Zagađeni/Iskrivljeni naziv atributa)	3.272	.040	2<1,3
Neprecizni podaci: Naziv aktivnosti (Istoimeni naziv atributa)	3.789	.024	2<1,3

* 1 – Istraživač; 2 – Praktikant; 3 – Oba.

Naredna analiza varijansi je primenjena kako bi se ispitala razlika u značaju koji ispitanici različitih uloga pridaju tehnikama čišćenja dnevnika izvršenja događaja prilikom pripreme podataka za dalju analizu. I ovaj test varijanse se pokazao kao statistički značajan, sa konkretnih tehnikama čišćenja u pogledu kojih postoji razlika između percepcije značajnosti od strane različitih uloga ispitanika prikazanim u tabeli 51. LSD post hok test je pokazao da praktikanti manje nego ostale dve grupe smatraju značajnim tehnike filtriranja putanje/događaja, abstrakcije događaja i klasterovanje putanje. Sa druge strane, istraživači smatraju značajnim manje nego praktikanti tehnike popravke dnevnika izvršenja događaja, a manje nego ostale dve grupe tehnike ugrađene obrade podataka.

Tabela 51. Razlike u značaju koji ispitanici različitih uloga pridaju tehnikama čišćenja dnevnika izvršenja događaja

Tehnike čišćenja dnevnika izvršenja događaja	F	Značajnost (p-vrednost)	Post hok LSD test*
Klasterovanje putanje	4.728	.010	2<1,3
Tehnike popravke dnevnika izvršenja događaja	5.093	.007	1<2
Filtriranje putanje/događaja	7.331	.001	2<1,3
Abstrakcija događaja	15.184	.000	2<1,3
Ugrađena obrada podataka	3.843	.023	1<2,3

* 1 – Istraživač; 2 – Praktikant; 3 – Oba.

4.2.4.2. Razlike u pogledu učestalosti susreta ispitanika različitih uloga sa problemima kvaliteta podataka i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja

Rezultat analize varijansi sprovedene kako bi se ispitala razlika u pogledu učestalosti susreta ispitanika različitih uloga sa problemima kvaliteta podataka pokazao je da postoji statistički značajna razlika između ispitanika u pogledu određenih problema, prikazanih u tabeli 52.

LSD post hok test je pokazao da se praktikanti manje nego druge dve grupe uloga susreću sa sledećim problemima kvaliteta dnevnika izvršenja događaja: nedostajući podaci o događajima; nedostajući podaci o vezi; nedostajući podaci o resursima i netačni podaci o vremenskoj odrednici. Takođe, LSD post hok test je pokazao da se istraživači više nego dve druge grupe uloga susreću sa problemima neprecizne vremenske odrednice, irelevantnih podataka o slučajevima procesa i irelevantnih podataka o događajima. LSD post hok test je pokazao i da se ispitanici koji su praktikanti i istraživači više nego praktikanti susreću sa problemom netačnih podataka o resursima, a susreću se više nego samo istraživači i praktikanti sa problemima nedostajućih podataka o resursima, netačnih podataka o slučaju, netačnih podataka o događajima, netačnih podataka o vezi i netačnih podataka o nazivima aktivnosti.

Tabela 52. Razlike u pogledu učestalosti susreta ispitanika različitih uloga sa problemima kvaliteta podataka

Problemi kvaliteta podataka	F	Značajnost (p-vrednost)	Post hok LSD test*
Nedostajući podaci: Događaj (Raštrkani događaj)	7.133	.001	2<1,3
Nedostajući podaci: Veza (Neuhvatljiv slučaj)	24.974	.000	2<1,3
Nedostajući podaci: Atribut slučaja/događaja	10.468	.000	3>1,2
Nedostajući podaci: Resurs	15.214	.000	3>1,2 2<1,3
Netačni podaci: Slučaj	11.751	.000	3>1,2
Netačni podaci: Događaj	9.049	.000	3>1,2
Netačni podaci: Veza (Raštrkani slučaj)	21.497	.000	3>1,2
Netačni podaci: Naziv aktivnosti (Zagađeni/Iskrivljeni naziv atributa)	3.811	.024	3>1,2
Netačni podaci: Vremenska odrednica (Hvatanje događaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj)	6.654	.002	2<1,3
Netačni podaci: Resurs (Zagađeni naziv atributa)	3.607	.029	3>2
Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj)	4.614	.011	1>2
Irelevantni podaci: Slučaj	6.237	.002	1>3
Irelevantni podaci: Događaj (Hvatanje događaja zasnovano na obrascu, Kolateralni događaj)	5.043	.007	1>2,3

* 1 – Istraživač; 2 – Praktikant; 3 – Oba.

Rezultat analize varijansi sprovedene kako bi se ispitala razlika u pogledu učestalosti primene tehnika čišćenja dnevnika izvršenja događaja od strane ispitanika različitih uloga pokazao je da postoji statistički značajna razlika. Tehnike čišćenja dnevnika izvršenja događaja kod kojih postoji statistički značajna razlika prikazani su u tabeli 53.

LSD post hok test je pokazao da praktikanti manje nego druge dve grupe uloga primenjuju klasterovanje putanja i filtriranje putanja/događaja, dok istraživači u većoj meri koriste tehnike popravke dnevnika izvršenja događaja i mašinsko učenje, veštačku inteligenciju i istraživanje i eksploataciju podataka. Takođe, ispitanici koji su ujedno i praktikanti i istraživači u manjoj meri u odnosu na druge dve grupe koriste abstrakciju događaja.

Tabela 53. Razlike u pogledu učestalosti primene tehnika čišćenja dnevnika izvršenja događaja od strane ispitanika različitih uloga

Tehnike čišćenja dnevnika izvršenja događaja	F	Značajnost (p-vrednost)	Post hok LSD test*
Klasterovanje putanja	8.136	.000	2<1,3
Tehnike popravke dnevnika izvršenja događaja	10.832	.000	1>2,3
Filtriranje putanje/događaja	10.161	.000	2<1,3
Abstrakcija događaja	3.616	.029	3>1,2
Mašinsko učenje, veštačka inteligencija i istraživanje i eksploatacija podataka	4.355	.014	1>2
Ugrađena obrada podataka	6.404	.002	2>1

* 1 – Istraživač; 2 – Praktikant; 3 – Oba.

4.2.4.3. Razlike u značaju koji ispitanici različitog nivoa iskustva pridaju problemima kvaliteta podataka dnevnika izvršenja događaja i tehnikama čišćenja dnevnika izvršenja događaja

Analiza varijansi je primenjena i na ispitivanje razlike u značaju koji ispitanici različitog nivoa iskustva pridaju problemima kvaliteta podataka dnevnika izvršenja događaja i tehnikama čišćenja dnevnika izvršenja događaja. Nivo iskustva ispitanika je izmeren putem Likertove skale, gde ispitanici mogu imati iskustvo manje od jedne godine, između 1 i 5 godina, između 6 i 10 godina i više od 10 godina. Analiza varijansi je pokazala statistički značajnu razliku između ispitanika različitog nivoa iskustva i u pogledu njihove percepcije značaja problema kvaliteta podataka i u pogledu značaja koji pridaju određenim tehnikama čišćenja dnevnika izvršenja događaja.

Tabela 54. prikazuje probleme kvaliteta kod kojih postoji statistički značajna razlika, kao i post hok LSD test razlike između ispitanika. Ispitanici sa iskustvom većim od 10 godina pridaju manje značaja nego oni sa iskustvom od manje od 5 godina problemu nedostajućih podataka o slučaju procesa. Ispitanici sa iskustvom između 1 i 5 godina pridaju manje značaja nego ispitanici sa iskustvom dužim od 6 godina problemima nedostajućih podataka o atributu slučaja/događaja i nedostajućih podataka o vremenskoj odrednici. Takođe, najmanje iskusni ispitanici sa iskustvom manjim od jedne godine pridaju manje značaja nego sve ostale grupe problem netačnih podataka o vezi između slučaja i događaja.

Tabela 54. Razlike u značaju koji ispitanici različitog nivoa iskustva pridaju problemima kvaliteta podataka dnevnika izvršenja događaja

Problemi kvaliteta podataka	F	Značajnost (p-vrednost)	Post hok LSD test*
Nedostajući podaci: Slučaj	2.905	0.036	4<1,2
Nedostajući podaci: Atribut slučaja/događaja	3.677	0.013	2<3,4
Nedostajući podaci: Vremenska odrednica	5.556	0.001	2<3,4
Netačni podaci: Veza (Raštrkani slučaj)	2.694	0.047	1<2,3,4

* 1 – manje od jedne godine; 2 – 1-5 godina; 3 – 6-10 godina; 4 – više od 10 godina.

Tabela 55. Prikazuje tehnike čišćenja dnevnika izvršenja događaja kod kojih postoji statistički značajna razlika u pogledu primene od strane ispitanika različitog iskustva. Post hok LSD test pokazuje da ispitanici sa iskustvom većim od 10 godina više primenjuju tehnike klasterovanja putanja i abstrakcije događaja nego sve ostale grupe ispitanika. Takođe, ispitanici sa iskustvom između 1 i 5 godina u većoj meri primenjuju tehnike popravke dnevnika izvršenja događaja.

Tabela 55. Razlike u značaju koji ispitanici različitog nivoa iskustva pridaju tehnikama čišćenja dnevnika izvršenja događaja

Tehnike čišćenja dnevnika izvršenja događaja	F	Značajnost (p-vrednost)	Post hok LSD test*
Klasterovanje putanja	2.62	0.052	4>1,2,3
Tehnike popravke dnevnika izvršenja događaja	4.292	0.006	2>1,3,4
Abstrakcija događaja	2.933	0.035	4>1,2,3

* 1 – manje od jedne godine; 2 – 1-5 godina; 3 – 6-10 godina; 4 – više od 10 godina.

4.2.5. Rezultat ispitivanja hipoteze istraživačkog modela II

Kako bi se ispitala hipoteza H6 istraživačkog modela 2 sprovedene su statističke analize Hi-kvadrat test i Pirsonov test korelacije. Na osnovu potvrđene hipoteze H6, sprovedena je i unakrsna analiza, kako bi se definisala uputstva za odabir tehnike čišćenja dnevnika izvršenja događaja na osnovu problema kvaliteta podataka.

4.2.5.1. Rezultat Pirsonovog testa korelacije između učestalosti susreta problema kvaliteta podataka i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja

Pirsonov test korelacije primenjen radi ispitivanja povezanosti između učestalosti susreta problema kvaliteta podataka i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja pokazao je da između posmatranih varijabli postoji statistički značajna korelacija. Tabela 56. pokazuje da za svaki problem kvaliteta podataka postoji bar jedna tehnika čišćenja dnevnika izvršenja događaja sa kojom je ona povezana. Statistički značajne korelacije su obeležene sa „*” i variraju od korelacije slabe do umerene jačine.

Tabela 56. Rezultati Pirsonovog testa korelacije – Učestalost susreta problema kvaliteta podataka i učestalost primene tehnika čišćenja dnevnika izvršenja događaja

Problemi kvaliteta podataka	Tehnike čišćenja dnevnika izvršenja događaja						
	Klasterovanje putanja	Tehnike popravke	Filtriranje putanje/događaja	Abstrakcija događaja	VI, MU, DU	Tehnike poravnavanja	Ugrađena obrada podataka
Nedostajući podaci: Slučaj	0.276*	0.378**	0.278*	0.240*	0.313*	0.102	0.057
Nedostajući podaci: Događaj (Raštrkani događaj)	0.023	0.144*	0.152*	0.182*	0.167*	0.047	0.058
Nedostajući podaci: Veza (Neuhvatljiv slučaj)	0.393*	0.131	0.228*	0.346*	0.157*	0.108	-0.104
Nedostajući podaci: Naziv aktivnosti	0.157*	0.299**	-0.123	0.162*	0.148*	0.222*	0.188*
Nedostajući podaci: Atribut slučaja/događaja	0.152*	-0.043	-0.092	0.198*	-0.054	0.044	0.104
Nedostajući podaci: Vremenska odrednica	0.046	0.125	-0.014	0.061	-0.107	-0.167*	-0.074
Nedostajući podaci: Resurs	0.101	-0.178*	0.048	0.179*	0.079	0.024	-0.036
Netačni podaci: Slučaj	0.145*	0.146*	0.197*	0.119	0.075	0.129	0.155*
Netačni podaci: Događaj	0.058	0.193**	0.052	0.019	0.005	-0.047	0.146*
Netačni podaci: Veza (Raštrkani slučaj)	0.177*	0.220**	0.043	0.195*	0.004	0.086	-0.022
Netačni podaci: Naziv aktivnosti (Zagađeni/Iskrivljeni naziv atributa)	0.083	0.147*	-0.155*	0.007	0.095	0.144*	0.309*
Netačni podaci: Atribut slučaja/događaja	0.222*	0.298**	-0.112	0.034	.275**	0.210*	0.123
Netačni podaci: Vremenska odrednica (Hvatanje događaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj)	0.053	0.222**	0.140*	0.116	-0.066	-0.093	-0.082
Netačni podaci: Resurs (Zagađeni naziv atributa)	0.099	0.193**	-0.075	0.157*	0.074	0.141*	0.157*

Neprecizni podaci: Veza	0.154*	0.288**	0.021	0.131	0.145*	0.179*	0.140*
Neprecizni podaci: Naziv aktivnosti (Istoimeni naziv atributa)	0.214* *	0.158*	0.019	0.08	0.180*	0.081	0.261* *
Neprecizni podaci: Atribut slučaja/događaja (Sinonimni naziv atributa)	0.224* *	0.152*	0.019	0.142*	0.161*	0.082	0.260* *
Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj)	0.072	0.175*	0.167*	0.123	-0.075	-0.001	0.012
Neprecizni podaci: Resurs	0.089	0.189**	-0.156*	0.155*	0.122	0.218* *	0.119
Irelevantni podaci: Slučaj	-0.008	-0.073	-0.089	-0.033	0.032	-0.02	0.158*
Irelevantni podaci: Događaj (Hvatanje događaja zasnovano na obrascu, Kolateralni događaj)	-0.093	-0.084	-0.01	-0.024	-0.021	-0.096	0.224* *
Količina podataka, nivo detaljnosti događaja, kompleksnost	0.218* *	0.098	0.247* *	0.236* *	0.144*	-0.09	0.108

** Korelacija je značajna na nivou od 0.01.

* Korelacija je značajna na nivou od 0.05.

4.2.5.2. Rezultat Hi-kvadrat testa povezanosti problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja

Kako bi se ispitala povezanost problema kvaliteta podataka dnevnika izvršenja događaja i tehnika čišćenja dnevnika izvršenja događaja, primenjen je Hi-kvadrat test. Tabela 57. predstavlja rezultate primene Hi-kvadrat testa koji ukazuje na statistički značajnu vezu između posmatranih varijabli (Hi-kvadrat = 1025.284, df 160, p = 0.000). Sve pretpostavke Hi-kvadrat testa su ispunjene, sa 11.2% ćelija sa očekivanim brojem manjim od 5.

Tabela 57. Rezultat ispitivanja povezanosti problema kvaliteta podataka dnevnika izvršenja događaja i tehnika čišćenja dnevnika izvršenja događaja

	Vrednost	df	p-vrednost
Hi-kvadrat	1025.284 ^a	160	0.000
Racio verovatnoće	1052.269	160	0.000
Broj validnih slučajeva	4242		

a. 21 ćelija (11.1%) ima očekivani broj manji od 5. Minimum očekivani broj je 0.62.

Dalje, izračunat je Kramerov koeficijent koji ukazuje na jačinu veze između problema kvaliteta podataka dnevnika izvršenja događaja i tehnika čišćenja dnevnika izvršenja događaja i prikazan je u tabeli 58. Veza između posmatranih varijabli se stoga tumači kao veza umerene jačine, sa Kramerovim koeficijentom od .174 i statističkom značajnošću testa od p = 0.000.

Tabela 58. Vrednost Kramerovog koeficijenta veze između problema kvaliteta podataka dnevnika izvršenja događaja i tehnika čišćenja dnevnika izvršenja događaja

Mere simetrije		Vrednost	p-vrednost
Nominalna po nominalnoj	Kramer V	0.174	0.000
Broj validnih slučajeva	4242		

S obzirom na prethodno prikazan Hi-kvadrat test koji potvrđuje povezanost umerene jačine između problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja, kao i na korelaciju dokazanu putem Pirsonovog testa korelacije, može se zaključiti da se **hipoteza 6 iz istraživačkog modela 2 prihvata** i da je moguće odrediti pogodnu tehniku čišćenja dnevnika izvršenja događaja u zavisnosti od problema kvaliteta podataka. Na osnovu toga, tabela 59. predstavlja tabelu kontingencije dobijenu unakrsnom analizom, koja pokazuje paterne odnosa između problema kvaliteta podataka i dnevnika izvršenja događaja. Tabela kontingencije u svakoj ćeliji sadži procentualno izražen udeo kojim ispitanici primenjuju određenu tehniku čišćenja dnevnika događaja u zavisnosti od problema kvaliteta podataka. Dodatna kategorija tehnika čišćenja dnevnika izvršenja događaja dodata u tabelu kontingencije na osnovu odgovora ispitanika je SQL.

Tabela 59. Tabela kontingencije odnosa između problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja u procentima

Problem kvaliteta podataka	Tehnike čišćenja dnevnika izvršenja događaja							SQL
	Tehnike poravnavanja	Ugrađena obrada podataka	Abstrakcija događaja	VI, MU, DU*	Tehnike popravke	Klasterovanje putanja	Filtriranje puatanje/dog.	
Nedostajući podaci: Slučaj	11.9	1.0	5.4	14.9	21.8	25.7	19.3	0
Nedostajući podaci: Događaj (Raštrkani događaj)	1.5	6.9	17.3	16.8	20.8	8.4	28.2	0
Nedostajući podaci: Veza (Neuhvatljiv slučaj)	3.2	5.0	21.2	13.8	12.1	26.3	18.4	0
Nedostajući podaci: Naziv aktivnosti	6.4	3.0	25.2	21.8	.5	21.3	10.9	0
Nedostajući podaci: Atribut slučaja/događaja	9.4	5.0	4.0	33.7	13.4	5.0	29.7	0
Nedostajući podaci: Vremenska odrednica	6.9	6.9	12.4	15.8	33.7	4.5	18.8	0
Nedostajući podaci: Resurs	5.0	6.4	5.4	25.7	3.5	9.9	21.8	3.5
Netačni podaci: Slučaj	6.4	6.9	2.0	4.5	28.2	19.3	27.7	5.0
Netačni podaci: Događaj	4.0	5.9	9.9	10.4	27.7	8.9	29.2	4.0
Netačni podaci: Veza (Raštrkani slučaj)	5.9	9.9	6.9	11.4	17.8	11.9	32.2	4
Netačni podaci: Naziv aktivnosti (Zagađeni/Iskrivljeni naziv atributa)	7.4	5.9	15.8	16.3	25.7	5.0	19.8	4
Netačni podaci: Atribut slučaja/događaja	4.0	9.4	13.4	19.3	22.3	5.9	24.8	0
Netačni podaci: Vremenska odrednica (Hvatanje događaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj)	11.9	5.9	6.9	10.9	30.7	4.0	28.7	0

Netačni podaci: Resurs (Zagađeni naziv atributa)	2.0	7.9	17.8	10.9	19.8	10.9	24.8	5.9
Neprecizni podaci: Veza	6.4	12.4	11.4	12.4	18.8	8.9	24.3	5.4
Neprecizni podaci: Naziv aktivnosti (Istoimeni naziv atributa)	5.0	11.4	30.7	8.9	16.3	5.0	17.8	5
Neprecizni podaci: Atribut slučaja/događaja (Sinonimni naziv atributa)	5.0	12.4	7.9	10.9	15.3	18.8	24.3	5.4
Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj)	2.5	9.4	6.9	14.9	33.2	9.9	17.8	4.5
Neprecizni podaci: Resurs	4.5	10.4	9.9	22.3	7.9	11.4	26.7	5.4
Irelevantni podaci: Slučaj	2.0	10.4	6.9	7.4	6.4	5.0	58.4	3.5
Irelevantni podaci: Događaj (Hvatanje događaja zasnovano na obrascu, Kolateralni događaj)	0	4.5	9.9	9.4	5.4	5.0	58.4	3.5
Količina podataka, nivo detaljnosti događaja, kompleksnost	4.0	6.9	15.3	16.8	1.0	10.4	42.1	3.5

* Veštačka inteligencija – VU; Mašinsko učenje – MU; Duboko učenje – DU.

Pirsonov test korelacije predstavljen u tabeli 56 i tabela kontingencije predstavljen u tabeli 59 prikazuju sledeće rezultate.

Prema 25.7% ispitanika, ukoliko je instance poslovnog procesa izvršena u stvarnosti, ali nije zabeležena u dnevniku izvršenja događaja, odnosno pojavi se problem nedostajućeg slučaja, tehnika iz grupe klasterovanja putanja bi trebala da bude primenjena. Problem nedostajućih podataka takođe ima statistički značajnu, slabu vezu sa tehnikama klasterovanja putanja. Takođe, prema 21.8% ispitanika, tehnike popravke dnevnika se mogu primeniti radi uklanjanja problema nedostajućeg slučaja. Veza između tehnika popravke dnevnika i problema nedostajućih podataka je statistički značajna i umerene jačine. Naredna tehnika koja se koristi od strane 19.3% ispitanika za uklanjanje problema nedostajućih podataka je tehnika filtriranja putanja/događaja dnevnika izvršenja događaja, sa dokazanom statistički značajnom vezom slabog intenziteta. Konačno, ispitanici su u 14.9% slučajeva za rešavanje ovog problema odabrali grupu tehnika veštačka inteligencija, mašinsko učenje i duboko učenje, koja sa problemom nedostajućih podataka o slučaju imaju statistički značajnu vezu umerene jačine. Ostale grupe tehnika su za rešavanje problema nedostajućih slučajeva odabrane u manje od 14.9% slučajeva.

Prilikom odabira tehnika za rešavanje problema nedostajućih podataka o događajima, odnosno prilikom pojave paterna raštrkanog događaja, ispitanici su birali tehnike filtriranja putanje/događaja u 28.2% slučajeva, tehnike popravke dnevnika u 20.8% slučajeva, abstrakciju događaja u 17.3% slučajeva i tehnike veštačke inteligencije, mašinskog učenja i dubokog učenja u 16.8% slučajeva. Sve pomenute tehnike imaju statistički značajnu, slabu vezu sa problemom nedostajućih događaja.

Problem nedostajućih podataka o vezi između slučaja procesa i događaja tog slučaja su ispitanici u najvećoj meri rešavali primenom klasterovanja putanja (26.3%), abstrakcijom događaja (21.2%) i filtriranjem putanja/događaja (18.4%). Abstrakcija događaja, klasterovanje putanja i filtriranje putanja/događaja imaju statistički značajnu, slabu vezu sa problemom nedostajućih podataka o vezi. Tehnike veštačke inteligencije, mašinskog učenja i dubokog učenja imaju statistički značajnu, slabu vezu sa problemom nedostajućih podataka o vezi, ali nisu često primenjivani za njegovo rešavanje od strane ispitanika.

Problem nedostajućih podataka o nazivu aktivnosti se od strane ispitanika najčešće rešava primenom tehnika abstrakcije događaja (25.2%), tehnika veštačke inteligencije, mašinskog učenja i

dubokog učenja (21.8%) i klasterovanja putanja (21.3%). Problem nedostajućih podataka o nazivu aktivnosti ima statistički značajnu vezu sa svim tehnikama čišćenja dnevnika izvršenja događaja, osim sa filtriranjem putanja/događaja.

Problem nedostajućih podataka o atributima slučaja procesa ili događaja ispitanici su u 33.7% slučajeva rešavali primenom tehnika veštačke inteligencije, mašinskog učenja i dubokog učenja. Takođe, 29.7% ispitanika je navelo da primenjuju tehniku filtriranja putanja/događaja kako bi uklonili slučajeve procesa ili događaje sa nedostajućim podacima. Problem nedostajućih podataka o atributima slučaja procesa ili događaja ima statistički značajnu vezu slabog intenziteta sa tehnikama abstrakcije događaja i klasterovanja putanja.

Problem nedostajućih podataka o vremenskoj odrednici 33.7% ispitanika rešava primenom tehnika popravke dnevnika izvršenja događaja. Dalje, 18.8% ispitanika je navelo da bi problem rešili sa tehnikom filtriranja putanja/događaja, dok bi 15.8% uklonilo problem nedostajućih vremenskih odrednica primenom tehnika veštačke inteligencije, mašinskog učenja i dubokog učenja. Prema Pirsonovom testu korelacije, nijedna tehnika čišćenja dnevnika izvršenja događaja nema pozitivnu statistički značajnu vezu sa ovim problemom, dok tehnike poravnavanja imaju statistički značajnu negativnu vezu slabog intenziteta.

Problem nedostajućih podataka o resursima 25.7% ispitanika uklanja primenom primenom tehnika veštačke inteligencije, mašinskog učenja i dubokog učenja, dok bi 21.8% ispitanika primenilo tehnike filtriranja putanja/događaja. Tehnika čišćenja dnevnika izvršenja događaja koja ima statističku značajnu, pozitivnu vezu sa ovim problemom je tehnika abstrakcije događaja, dok tehnika popravke dnevnika ima statistički značajnu, negativnu vezu sa problemom nedostajućih podataka o resursima.

Problem netačnih podataka o slučaju procesa 28.2% ispitanika rešava primenom tehnikak popravke dnevnika, 27.7% ispitanika rešava primenom tehnika filtriranja putanja/događaja, dok 19.3% ispitanika rešava primenom tehnika klasterovanja putanja procesa. Ove tehnike čišćenja dnevnika izvršenja događaja imaju statistički značajne, slabe veze sa problemom netačnih podataka o slučaju procesa.

Problem netačnih podataka o događajima procesa 29.2% ispitanika bi uklonilo primenom tehnika filtriranja putanja/događaja, dok bi 27.7% ispitanika primenilo tehnike popravke dnevnika izvršenja događaja. Tehnike čišćenja dnevnika izvršenja događaja koje imaju statističku značajnu, pozitivnu vezu sa ovim problemom su tehnike popravke dnevnika i tehnika abstrakcije događaja.

Problem netačnih podataka o vezi između slučaja i događaja procesa 32.2% ispitanika bi uklonilo primenom tehnika filtriranja putanja/događaja, dok bi 17.8% ispitanika primenilo tehnike popravke dnevnika izvršenja događaja. Tehnike čišćenja dnevnika izvršenja događaja koje imaju statističku značajnu, pozitivnu vezu sa ovim problemom su tehnike popravke dnevnika i tehnika abstrakcije događaja. Tehnike popravke dnevnika i ugrađena obrada podataka imaju statistički značajnu, pozitivnu vezu sa problemom netačnih podataka o vezi između slučaja i događaja procesa.

Problem netačnih podataka o nazivima aktivnosti 25.7% ispitanika bi uklonilo primenom tehnika filtriranja putanja/događaja, dok bi 19.8% ispitanika primenilo tehnike popravke dnevnika izvršenja događaja. Dalje, 16.3% ispitanika bi primenilo tehnike veštačke inteligencije, mašinskog učenja i dubokog učenja, a 15.8% ispitanika bi primenilo abstrakciju događaja. Tehnike čišćenja dnevnika izvršenja događaja koje imaju statističku značajnu, pozitivnu vezu, slabog intenziteta sa ovim problemom su tehnike popravke dnevnika i tehnike poravnavanja. Tehnika filtriranja putanja/događaja ima statističku značajnu, negativnu vezu slabog intenziteta, dok tehnika ugrađene obrade podataka ima statistički značajnu, pozitivnu vezu umerenog intenziteta.

Problem netačnih podataka o atributima slučaja i događaja, vremenskim odrednicama i resursima ispitanici najčešće rešavaju primenom tehnika popravke dnevnika izvršenja i filtriranja putanja/događaja. Takođe, svi problemi netačnih podataka imaju statistički značajnu, slabu do umerenu vezu sa tehnikama popravke dnevnika izvršenja događaja.

Problem nepreciznih podataka o vezi između slučaja procesa i događaja koji mu pripadaju ispitanici u 24.3% slučajeva rešavaju primenom tehnika filtriranja putanja/događaja, dok u 18.8%

slučajeva primenjuju tehnike popravke dnevnika. Tehnike čišćenja dnevnika izvršenja događaja koje imaju statističku značajnu, pozitivnu vezu sa ovim problemom su tehnike popravke dnevnika, ugrađene obrade podataka, klasterovanja putanja/događaja, tehnike poravnavanja i tehnike veštačke inteligencije, mašinskog učenja i dubokog učenja.

Problem nepreciznih podataka o nazivu aktivnosti se od strane ispitanika najčešće rešava primenom tehnika abstrakcije događaja (30.7%), tehnika filtriranja putanja/događaja (17.8%) i popravke dnevnika (16.3%). Problem nepreciznih podataka o nazivu aktivnosti ima statistički značajnu vezu sa tehnikama veštačke inteligencije, mašinskog učenja i dubokog učenja, tehnikama popravke dnevnika, klasterovanjem putanja i ugrađenom obradom podataka.

Problem nepreciznih podataka o atributima slučaja i događaja ispitanici u 24.3% slučajeva koriste tehnike filtriranja putanja/događaja, u 18.8% slučajeva koriste tehniku klasterovanje putanja, a u 15.3% slučajeva tehnike popravke dnevnika. Problem nepreciznih podataka o atributima slučaja i događaja ima statistički značajnu vezu sa tehnikama veštačke inteligencije, mašinskog učenja i dubokog učenja, tehnikama popravke dnevnika, klasterovanjem putanja i ugrađenom obradom podataka.

Problem nepreciznih podataka o vremenskoj odrednici ispitanici u 33.2% slučajeva rešavaju primenom tehnika popravke dnevnika, dok u 17.8% slučajeva problem rešavaju primenom tehnika filtriranja putanje/događaja. Problem nepreciznih podataka o atributima slučaja i događaja ima statistički značajnu, slabu vezu sa navedenim tehnikama.

Problem nepreciznih podataka o podacima resursa procesa ispitanici u 26.7% slučajeva rešavaju primenom filtriranja putanje/događaja, dok u 22.3% slučajeva koriste tehnike veštačke inteligencije, mašinskog učenja i dubokog učenja. Problem nepreciznih podataka o podacima resursa procesa ima statistički značajnu vezu sa tehnikama popravke dnevnika, filtriranja putanje/događaja, abstrakcije događaja i tehnikama poravnavanja.

Više od 50% ispitanika bi probleme irelevantnih podataka rešilo primenom tehnika filtriranja putanje/događaja, a 42% ispitanika bi primenilo istu tehniku radi ukljanjanja problema velike količine podataka, nivoa detaljnosti događaja i kompleksnosti dnevnika izvršenja događaja. Problem irelevantnih podataka slučaja i događaja ima statistički značajnu, pozitivnu vezu sa tehnikom ugrađene obrade podataka.

5. DISKUSIJA REZULTATA ISTRAŽIVANJA

U okviru ovog poglavlja predstavljena je analiza izvođenja istraživanja i tumačenje rezultata istraživanja u kontekstu značaja rezultata u oblasti automatskog istraživanja poslovnih procesa i teorijskih i praktičnih implikacija doktorske disertacije. Poglavlje je podeljeno na dva dela kako bi se predstavila diskusija rezultata dva konkurentna istraživačka modela, koji odgovaraju na zasebna istraživačka pitanja.

5.1. Diskusija rezultata istraživačkog modela I

Prvi deo doktorske disertacije je težio da odgovori na istraživačko pitanje: Koji faktori projekata analize i automatskog otkrivanja poslovnih procesa (projektna karakteristike) utiču na odabir tipova, softverskih alata i algoritama/tehnika automatskog otkrivanja poslovnih procesa?

Kako bi se postavila pretpostavke o ključnim faktorima projekata automatskog otkrivanja poslovnih procesa i njihovim potencijalnim vezama, izvršeno je nekoliko sistematskih pregleda literature. Prvi sistematski pregled literature je imao za cilj da izvrši eksploratornu analizu primarnih studija iz oblasti automatskog otkrivanja poslovnih procesa, kako bi se utvrdilo koje faktore projekata automatskog otkrivanja poslovnih procesa istraživači i autori smatraju značajnim. Rezultat navedenog sistematskog pregleda literature, prikazanog u 2.2.1, je pokazao da su potencijalni ključni faktori istraživanja tipovi i perspektive AOPP-a, grane industrije i vrste poslovnih procesa koji su analizirani, kao i primenjene tehnike automatskog otkrivanja poslovnih procesa unutar odgovarajućih softverskih alata. Zaključci predstavljenog pregleda literature o primeni automatskog otkrivanja poslovnih procesa u industriji su pokazali da se tipovi i perspektive ne primenjuju jednako u svakoj industriji, kao i da se vrste analiziranih poslovnih procesa razlikuju. Samim tim, u različitim scenarijima su primenjeni različiti tipovi, algoritmi, tehnike i softverski alati.

Kako bi se dalje ispitao odnos između ključnih faktora primene automatskog otkrivanja poslovnih procesa, sproveden je sistematski pregled literature samo u proizvodnoj industriji, gde su primenjene naprednije statističke metode kako bi se ispitala njihova povezanost i dodata je varijabla cilj analize. Rezultati navedenog sistematskog pregleda literature su predstavljeni u potpoglavlju 2.2.2., sa odgovorima na istraživačka pitanja i zaključkom da, na posmatranom uzorku, postoji potencijalna veza između ciljeva projekata i sprovedenih tipova i perspektiva AOPP-a.

Na osnovu predstavljenih teorijskih osnova, definisan je istraživački model 1 (Slika 10. Istraživački model 1), sa postavljenih 5 hipoteza. Na osnovu protokola pregleda literature, definisanog u skladu sa smernicama za izvođenje sistematskih pregleda literature, prikupljene su primarne studije koje su pružile podatke za ispitivanje hipoteza istraživačkog modela 1.

5.1.1. Diskusija procedure prikupljanja podataka i deskriptivne statistika varijabli istraživačkog modela I

Poglavlje 3.1.3. predstavlja bibliografske podatke primarnih studija uključenih u istraživanje, dok je spisak primarnih studija priložen kao Prilog A. Može se zaključiti da su ključni razlozi za odbacivanje radova bili sledeći kriterijumi odbacivanja i prihvatanja primarnih studija:

- Značajan broj radova se nije bavio primenom automatskog otkrivanja poslovnih procesa, a termin „*process mining*“ je bio uključen u njihov naslov ili abstrakt. Zaključuje se da prilikom primene tehnika iz oblasti istraživanja i eksploatacije podataka u kontekstu analize procesa, istraživači navode termin „*process mining*“, koji se u tom kontekstu ne odnosi na automatsko otkrivanje poslovnih procesa.
- Velik broj radova prikazuje primenu nove metodologije ili pristupa u oblasti automatskog otkrivanja poslovnih procesa, sa primenom nad veštački generisanim podacima dnevnika izvršenja događaja. Za potrebe ovog istraživanja, u obzir su uzeti isključivo podaci nastali izvršavanjem poslovnih procesa u stvarnim poslovnim sistemima.

- Velik broj radova je odbačen jer autori nisu precizno naveli podatke o varijablama posmatranim u ovom istraživanju, poput primenjenih tehnika i algoritama ili korišćenih softverskih alata.
- Indeksne baze u kojima je izvršena pretraga radova u velikoj meri sadrže iste radove, te je konačni broj primarnih studija iz tog razloga umanjen.

Potpoglavlje 4.1.1. sadrži rezultate deskriptivne statistike varijabli istraživačkog modela 1.

U pogledu rezultata analize učestalosti vrste industrije u okviru koje je primenjeno automatsko otkrivanje poslovnih procesa, rezultati delimično prikazuju očekivane podatke.

Zdravstvena industrija je vodeća industrija u pogledu primene različitih oblasti obrade podataka, koje teže da otkriju značajne uvide u poslovanje nekog zdravstvenog sistema. Samim tim, očekivan je podatak da je u oblasti automatskog otkrivanja poslovnih procesa zdravstvena industrija na vodećem mestu. Ukoliko se pogleda tabela 60, može se zaključiti da je i u periodu između 2009. i 2018. godine zdravstvena industrija zauzimala najveći udeo u primeni automatskog otkrivanja poslovnih procesa.

Proizvodna industrija je zauzela drugo mesto kao istaknuta industrija u kojoj se primenjuje automatsko otkrivanje poslovnih procesa. Ukoliko se taj podatak uporedi sa podatkom iz perioda između 2009. i 2018. godine, gde je proizvodna industrija zauzimala tek peto mesto, može se zaključiti da je postignut izuzetan napredak u ovom polju. Razlog za takav napredak leži u razvoju Industrije 5.0, čiji je fokus na saradnji između čoveka i mašina. Sa povećanjem stope automatizacije proizvodnih sistema, povećao se i broj potencijalnih izvora podataka za formiranje dnevnika izvršenja događaja i samim tim i mogućnosti primene automatskog otkrivanja poslovnih procesa.

Još jedna industrija u okviru koje je primećen veliki napredak jeste obrazovna industrija, čiji je udeo porastao sa 8.0% na 15.3%. Razlog za to leži u povećanoj primeni platformi za udaljeno učenje, uzrokovano pandemijom koja je trajala od marta 2020. godine do marta 2023. godine. Povećana učestalost primene platformi za udaljeno učenje je povećala i količinu podataka koja je beležena o procesima držanja nastave, učenja studenata i polaganja onlajn testova. Samim tim, povećava se i mogućnost primene automatskog otkrivanja poslovnih procesa.

IKT industrija je na četvrtom mestu u odnosu na ostale industrije. Udeo njene zastupljenosti se nije značajno promenio u odnosu na prethodni period, s obzirom da se proces razvoja softvera nije u značajnoj meri promenio. Takođe, može se zaključiti da je interesovanje istraživača, uslovljeno trendovima na tržištu, u pogledu ove industrije ostalo nepromenjeno. Preostale industrije se nisu pokazale od velikog značaja za primenu automatskog otkrivanja poslovnih procesa, sa stabilno niskim udelom u odnosu na ostale industrije.

Tabela 60. Poređenje zastupljenosti primene automatskog otkrivanja poslovnih procesa u različitim industrijama i vremenskim periodima

Industrija	Udeo u % 2019 - 2023	Udeo u % 2009 - 2018
Zdravstvo	33.3	28.0
Proizvodnja	18.8	8.0
Obrazovanje	15.3	8.0
Informaciono-komunikacione tehnologije	13.9	14.0
Javni servisi i usluge	5.6	9.0
Računovodstvo, kontroling, bankarstvo i osiguranje	4.9	11.0
Logistika	4.2	3.0
Građevina	3.5	0
Energija	0.7	8
Ukupno	100.0	100.0

U pogledu rezultata analize učestalosti postavljenih ciljeva analize, može se zaključiti da je otkrivanje modela procesa i unapređenje procesa značajno zastupljeniji u odnosu na ostale ciljeve. Ovakav ishod je očekivan, s obzirom da ovaj cilj obuhvata dva glavna tipa automatskog otkrivanja poslovnih procesa. U periodu od 2009. do 2021. godine, taj udeo je bio umereno manje značajan, sa udelom od 53.0% u odnosu na udeo od 66.0%. Bitno je istaći da su ciljevi predviđanja ponašanja procesa i pronalazak anomalija i nepredviđenih okolnosti, iako značajni u kontekstu analize u svim industrijama, veoma značajni u proizvodnoj industriji. Može se zaključiti da je razlog za to pogodnost samih proizvodnih sistema za sprovođenje takve vrste analize, gde je npr. proces proizvodnje visoko standardizovan i automatizovan proces i čiji su informacioni sistemi i razvijeni sa ciljem uočavanja odstupanja od predviđenog ponašanja. Ciljevi koje se nisu pominjali u proizvodnoj industriji u periodu od 2009. do 2021. godine, a zastupljeni u svim industrijama u periodu od 2019. do 2023. godine, su poređenje poslovnih procesa, analiza vremena čekanja, otkrivanje uzroka problema i otkrivanje poslovnih pravila. Ovi ciljevi su vrlo specifični i pokazuju da su istraživači upoznati sa svim mogućnostima koje pruža automatsko otkrivanje poslovnih procesa i da počinju da se udaljavaju od primene isključivo standardnih tehnika. Automatizacija poslovnog procesa i analiza resursa su takođe veoma specifični ciljevi, koji pokazuju da se oblast automatskog otkrivanja poslovnih procesa kreće u dobrom pravcu, ka primenama od veliko značaja za industrijske sisteme.

Tabela 61. Poređenje zastupljenosti postavljenih ciljeva analize automatskog otkrivanja poslovnih procesa u različitim vremenskim periodima

Cilj analize	Udeo u % Sve industrije 2019 - 2023	Udeo u % Proizvodna industrija 2009 - 2021
Otkrivanje i unapređenje poslovnog procesa	66.0	53.0
Predviđanje ponašanja procesa	9.0	20.0
Pronalazak anomalija i nepredviđenih aktivnosti	9.0	15.0
Poređenje poslovnih procesa	4.2	0
Analiza vremena čekanja	3.5	0
Analiza promena procesa	2.1	0
Automatizacija poslovnog procesa	2.1	2.5
Analiza resursa	2.1	2.5
Otkrivanje uzroka problema	1.4	0
Otkrivanje poslovnih pravila	0.7	0
Ukupno	100.0	100.0

Velika raznovrsnost poslovnih procesa, sa izdvojenih 42 različita poslovna procesa u Tabela 23, pokazuje da automatsko otkrivanje poslovnih procesa ima široku mogućnost primene. Iako su navedeni poslovni procesi vrlo različiti, neki od njih se mogu posmatrati kroz industrije u kojima su analizirani. Na primer, u okviru zdravstvene industrije, dominantam poslovni proces je putanja pacijenta, jer on nosi najveći udeo troškova koji se javljaju prilikom lečenja pacijenata. Unapređenjem ovog poslovnog procesa, zdravstveni sistemi mogu da steknu uvid rada zadržavstvenih radnika i da uoče aktivnosti koje povećavaju nezadovoljstvo pacijenata i smanjuju efikasnost rada zapsolenih. Takođe, dnevnik izvršenja događaja za putanju pacijenta je moguće lako kreirati ekstrakcijom podataka iz informacionog sistema bolnice. U proizvodnoj industriji, proces od najvećeg značaja je proizvodni proces, takođe iz razloga umanjavanja troškova proizvodnje unapređenjem performansi proizvodnog procesa i umanjjenjem devijacija i nepredviđenih okolnosti. Proces učenja i proces izrade softvera su takođe dominantni u svojim industrijama, odnosno u obrazovnoj i IKT industriji, iz razloga što predstavljaju ključne poslovne procese. Poslovni proces koji se analizirao u velikom broju primarnih studija, a nije usko vezan za specifičnu industriju, je proces ponašanja korisnika nekog informacionog sistema. Značaj ovog

poslovnog procesa proizilazi iz potrebe da se razume korišćenje informacionih tehnologija u svim industrijama kako bi se sami informacioni sistemi prilagodili korisnicima.

U pogledu poređenja zastupljenosti primene tipova automatskog otkrivanja poslovnih procesa u različitim vremenskim periodima predstavljenim u tabeli 62, može se zaključiti da su dva osnovna tipa automatskog otkrivanja poslovnih procesa zastupljena u jednakoj meri. Otkrivanje poslovnog procesa je očekivano najznačajniji tip, s obzirom da on podrazumeva osnovnu namenu automatskog otkrivanja poslovnih procesa. Unapređenje poslovnih procesa je uvek usko povezano sa otkrivanjem samog modela procesa iz razloga što većina softverskih alata ima tehnike koje računaju metrike oba tipa automatskog otkrivanja poslovnih procesa u isto vreme. Takođe, vrlo je jednostavno proširiti model procesa sa metrikama performansi procesa kako bi se izvršilo i njegovo unapređenje. Provera usaglašenosti je tip poslovnog procesa koji se često izvršava nezavisno od prva dva tipa automatskog otkrivanja poslovnih procesa. Takođe, može se zaključiti da se ovaj tip u manjoj meri izvršava u periodu između 2019. i 2023 godine nego u prethodnom periodu. Analiza društvenih mreža nema velik značaj u oblasti, iz razloga što postoje i druge oblasti koje podržavaju ovakav vid analize i istraživači ne postavljaju ciljeve koji zahtevaju analizu društvenih mreža. Ono što je veoma značajno je uvođenje prediktivnog automatskog otkrivanja poslovnih procesa kao novog tipa automatskog otkrivanja poslovnih procesa, što je u skladu sa najnovijom literaturom iz oblasti i ukazuje na rastući značaj predikcije u analizi poslovnih procesa.

Tabela 62. Poređenje zastupljenosti primene tipova automatskog otkrivanja poslovnih procesa u različitim vremenskim periodima

Tip AOPP	Udeo u % 2019 - 2023	Udeo u % 2009 - 2018
Otkrivanje poslovnog procesa	95.1	100.0
Unapređenje poslovnog procesa	76.4	60.0
Provera usaglašenosti	18.7	42.0
Analiza društvenih mreža	9.0	0
Prediktivno AOPP	5.5	0

Posmatranjem rezultata deskriptivne statistike u pogledu softverskih alata, algoritama i tehnika, vrlo jasno se uočavaju tendencije u njihovoj primeni. U pogledu otkrivanja poslovnih procesa i unapređenja poslovnih procesa, istraživači skoro isključivo primenjuju softverske alata *ProM*, *Disco* i *Celonis*, kao i algoritme fazi algoritam, induktivni algoritam, heuristički algoritam i ugrađene tehnike alata *Celonis*. U poređenju sa istraživanjem primene automatskog otkrivanja poslovnih procesa u prethodnom periodu (2009 - 2018. godine), može se zaključiti da su fazi algoritam, induktivni algoritam i heuristički algoritam, kao i alati *ProM* i *Disco* i dalje najznačajniji. Međutim, u periodu od 2019. do 2023. godine značajan udeo na tržištu uzima i *Celonis*. *Celonis* je komercijalni softverski alat koji se koristi u primeni u velikim industrijskim sistemima i koji ukazuje na širenje i zrelost automatskog otkrivanja poslovnih procesa kao oblasti obrade podataka. Analizirane pridružene tehnologije pokazuju da je zarad kompletnije primene i postizanja specifičnih ciljeva analize automatskog otkrivanja poslovnih procesa korisno primeniti i tehnike mašinskog učenja i istraživanja i eksploatacije podataka.

5.1.2. Diskusija rezultata ispitivanja hipoteza istraživačkog modela I

Istraživački model 1 sadrži ključne faktore projekata automatskog otkrivanja poslovnih procesa i hipoteze koje prepostavljaju postojanje veza između tih faktora odnosno varijabli istraživanja. U nastavku je dat komentar na rezultate ispitanih hipoteza predstavljenih u 4.1.2.

H1: Vrsta industrije ima povezanost sa vrstom poslovnog procesa koji se analizira putem automatskog otkrivanja poslovnih procesa

Navedena hipoteza H1 je ispitana primenom Fišerovog egzaktnog testa i na osnovu rezultata testa je prihvaćena. Stoga, može se zaključiti da određeni poslovni procesi koji su analizirani u okviru primarnih studija istraživačkog modela 1 pripadaju određenim industrijskim sistemima. Odnosno, u

okviru svakog posmatranog industrijskog sistema, postoje poslovni procesi koji mu pripadaju. Potrebno je naglasiti da u prethodnim istraživanjima ne postoje rezultati ispitivanja veze između vrste industrije i poslovnih procesa.

U pogledu istraživanja ove doktorske disertacije, prihvatanje hipoteze H1 sugerise da je industrija ključni faktor projekata automatskog otkrivanja poslovnih procesa, koji treba da bude uzet u obzir prilikom planiranja i izvođenja automatskog otkrivanja poslovnih procesa. U širem kontekstu, potvrđivanje povezanosti između vrste industrije i vrste poslovnih procesa upućuje na nekoliko praktičnih implikacija. Pre svega, menadžment organizacija koje se bave analizom poslovnih procesa kako bi unapredile poslovanje, može da donosi strateške odluke na osnovu informacije o tome koji poslovni procesi du od velikog značaja za određenu industriju. Takođe, ukoliko se utvrde ključni poslovni procesi neke industrije, moguće je alocirati resurse kako bi se optimizovali poslovni procesi koji stvaraju dodatnu vrednost organizaciji.

H2: Vrsta industrije ima povezanost sa ciljem istraživanja

Navedena hipoteza H2 je ispitana primenom Fišerovog egzaktnog testa i na osnovu rezultata testa je prihvaćena. Stoga, može se zaključiti da su određeni ciljevi istraživanja povezani sa vrstom industrije u okviru koje su postavljeni. Potrebno je naglasiti da u prethodnim istraživanjima ne postoje rezultati ispitivanja veze između vrste industrije i cilja istraživanja u nekoj oblasti.

U pogledu istraživanja ove doktorske disertacije, prihvatanje hipoteze H2 sugerise da je cilj analize ključni faktor projekata automatskog otkrivanja poslovnih procesa, koji treba da bude uzet u obzir prilikom planiranja i izvođenja automatskog otkrivanja poslovnih procesa. Takođe, stiču se uslovi za definisanje smernica za odabir tipa automatskog otkrivanja poslovnih procesa, na osnovu industrije u kojoj se istraživanje odvija i postavljenog cilja istraživanja. U širem kontekstu, poravnavanje projekata analize poslovnih procesa unutar određene industrije sa ciljevima analize koji su određeni kao ključni za tu industriju donosi nekoliko praktičnih implikacija za menadžere projekata, poput povećanog značaja i vrednosti analize. Takođe, utvrđivanje veze između industrije i cilja analize može da pomogne prilikom procesa prikupljanja i prioritizacije podataka za analizu poslovnih procesa.

H3: Cilj analize automatskog otkrivanja poslovnih procesa ima povezanost sa vrstom poslovnog procesa koji se analizira

Navedena hipoteza H3 je ispitana primenom Fišerovog egzaktnog testa i na osnovu rezultata testa je **odbaćena**. Stoga, može se zaključiti da ne postoji povezanost između cilja analize i vrste poslovnog procesa koji se analizira. Odnosno, različiti poslovni procesi mogu biti analizirani sa različitim ciljevima analize. Razlog za nepostojanje veze između ove dve varijable može biti u tome što su identifikovana 42 različita poslovna procesa, gde se velik broj poslovnih procesa pojavljuje samo jednom. Potrebno je naglasiti da u prethodnim istraživanjima ne postoje rezultati ispitivanja veze između cilja analize i vrste poslovnog procesa. S obzirom na odbacivanje hipoteze H3, vrsta poslovnog procesa nije uvrštena u smernice za odabir tipa automatskog otkrivanja poslovnih procesa u zavisnosti od scenarija u kojem se analiza odvija.

H4: Cilj analize automatskog otkrivanja poslovnih procesa ima povezanost sa tipom automatskog otkrivanja poslovnih procesa

Navedena hipoteza H4 je ispitana primenom Hi-kvadrat testa i na osnovu rezultata testa je **prihvaćena**. Stoga, može se zaključiti da su određeni ciljevi istraživanja povezani sa tipom automatskog otkrivanja poslovnih procesa koji je primenjen. Potrebno je naglasiti da u prethodnim istraživanjima ne postoje rezultati ispitivanja veze između cilja istraživanja i primenjenog tipa automatskog otkrivanja poslovnih procesa.

U pogledu istraživanja ove doktorske disertacije, prihvatanje hipoteze H4 sugerise da je tip automatskog otkrivanja poslovnih procesa takođe ključni faktor projekata automatskog otkrivanja poslovnih procesa i time se stiču uslovi za definisanje smernica za odabir tipa automatskog otkrivanja poslovnih procesa, na osnovu industrije u kojoj se istraživanje odvija i postavljenog cilja istraživanja. U širem kontekstu, dokazivanje postojanja veze između cilja analize i tipa automatskog otkrivanja poslovnog procesa podrazumeva nekoliko praktičnih implikacija za istraživače i praktikante, kao i za

menadžere projekata automatskog otkrivanja poslovnih procesa. Naime, osigurava se da je pristup analizi poslovnih procesa usklađen sa ciljom analize i omogućava se blagovremena alokacija resursa potrebnih za izvođenje određenog tipa automatskog otkrivanja poslovnih procesa.

H5: Cilj analize automatskog otkrivanja poslovnih procesa ima povezanost sa primenjenim algoritmima i tehnikama

Kako bi se ispitala hipoteza H5, bilo je potrebno sprovesti nekoliko nezavisnih Fišerovih egzaktnih testova, kao bi se ispitale sledeće povezanosti:

- cilj analize ↔ algoritmi za otkrivanje poslovnih procesa;
- cilj analize ↔ tehnike unapređenja poslovnih procesa;
- cilj analize ↔ tehnike provere usaglašenosti procesa;
- cilj analize ↔ tehnike analize društvenih mreža i
- cilj analize ↔ tehnike predviđanja.

Na osnovu potvrđenih navedenih veza, hipoteza H5 je **prihvaćena**. Može se zaključiti da cilj analize utiče na odluke praktikanata automatskog otkrivanja poslovnih procesa prilikom odabira algoritma ili tehnike koje će primeniti u zavisnosti od tipa automatskog otkrivanja poslovnih procesa koji se sprovodi. Međutim, s obzirom da je deskriptivnom analizom varijabli istraživačkog modela ustanovljeno da kod svakog tipa automatskog otkrivanja poslovnih procesa postoji mali broj dominantnih algoritama i tehnika koje se koriste, one neće biti direktno uključene u tabelu sa smernicama za odabir načina izvođenja automatskog otkrivanja poslovnih procesa, već kao zasebna tabela sa smernicama za odabir algoritma i tehnike automatskog otkrivanja poslovnih procesa u zavisnosti od tipa automatskog otkrivanja poslovnih procesa koji se primenjuje, nastale na osnovu deskriptivne statistike posmatranih varijabli primarnih studija. Dominantni algoritmi i tehnike su, će stoga biti naglašeni kao dodatak samim smernicama.

Potrebno je naglasiti da ne postoji prethodno istraživanje koje se bavi analizom veze između cilja analize i primenjenih algoritama i tehnika automatskog otkrivanja poslovnih procesa.

5.1.3. Smernice za planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa u zavisnosti od vrste industrije i cilja analize

Na osnovu prihvaćenih hipoteza o povezanosti vrste industrije sa ciljem analize i cilja analize sa tipovima automatskog otkrivanja poslovnih procesa i odgovarajućim algoritmima i tehnikama, kao i na osnovu tabele kontingencije odnosa između vrste proizvodnje, cilja analize i tipa automatskog otkrivanja poslovnih procesa, definisane su smernice za planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa. Scenario u okviru kojeg se izvršava projekat analize poslovnog procesa definisan je u odnosu na industriju i cilj analize automatskog otkrivanja poslovnih procesa. U zavisnosti od scenarija, tabela 63. prikazuje rangirane tipove automatskog otkrivanja poslovnih procesa, koji se mogu uspešno primeniti kako bi se ostvario cilj analize automatskog otkrivanja poslovnih procesa, gde rang 1 predstavlja tehniku najvišeg prioriteta određenog na osnovu procenta iz Tabela 59.

Tabela 63. Smernice za planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa u zavisnosti od vrste industrije i cilja analize

Industrija	Cilj analize	Tipovi automatskog otkrivanja poslovnih procesa				
		Rang				
		1	2	3	4	5
Proizvodnja	Otkrivanje i unapređenje poslovnog procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	Prediktivno AOPP	Analiza društvenih mreža
	Pronalazak anomalija i nepredviđenih aktivnosti	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	/	/

	Analiza vremena čekanja	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	/	/	/
	Predviđanje ponašanja procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Prediktivno AOPP	Analiza društvenih mreža	/
	Automatizacija poslovnog procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	/	/	/
	Otkrivanje poslovnih pravila	Unapređenje poslovnog procesa	/	/	/	/
	Analiza promena procesa	Otkrivanje poslovnog procesa	/	/	/	/
Zdravstvo	Otkrivanje i unapređenje poslovnog procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	Analiza društvenih mreža	/
	Poređenje poslovnih procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	/	/
	Predviđanje ponašanja procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Prediktivno AOPP	/	/
	Analiza resursa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Analiza društvenih mreža	/	/
	Analiza vremena čekanja	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	/	/	/
	Analiza promena procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	/	/	/
Obrazovanje	Otkrivanje i unapređenje poslovnog procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	/	/
	Otkrivanje uzroka problema	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	/	/	/
	Predviđanje ponašanja procesa	Otkrivanje poslovnog procesa	Provera usaglašenosti	/	/	/
	Pronalazak anomalija i nepredviđenih aktivnosti	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	/	/	/
	Automatizacija poslovnog procesa	Otkrivanje poslovnog procesa	/	/	/	/
	Poređenje poslovnih procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	/	/	/

Računovodstvo, kontroling, bankarstvo i osiguranje	Otkrivanje i unapređenje poslovnog procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Analiza društvenih mreža	/	/
	Predviđanje ponašanja procesa	Otkrivanje poslovnog procesa	Prediktivno AOPP	/	/	/
	Pronalazak anomalija i nepredviđenih aktivnosti	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	/	/	/
	Automatizacija poslovnog procesa	Otkrivanje poslovnog procesa	/	/	/	/
	Poređenje poslovnih procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	/	/
Informaciono- komunikacione tehnologije	Otkrivanje i unapređenje poslovnog procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	Analiza društvenih mreža	/
	Pronalazak anomalija i nepredviđenih aktivnosti	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	/	/
	Predviđanje ponašanja procesa	Otkrivanje poslovnog procesa	/	/	/	/
Logistika	Otkrivanje i unapređenje poslovnog procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	0	0	0
	Analiza promena procesa	Otkrivanje poslovnog procesa	Provera usaglašenosti	/	/	/
	Predviđanje ponašanja procesa	Prediktivno AOPP	/	/	/	/
	Analiza resursa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	/	/
Javni servisi i usluge	Otkrivanje i unapređenje poslovnog procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	/	/
	Otkrivanje uzroka problema	Unapređenje poslovnog procesa	/	/	/	/
	Pronalazak anomalija i nepredviđenih aktivnosti	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Analiza društvenih mreža	/	/
Građevina	Otkrivanje i unapređenje poslovnog procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Analiza društvenih mreža	Provera usaglašenosti	/
	Predviđanje ponašanja procesa	Prediktivno AOP	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	/	/

Energija	Otkrivanje i unapređenje poslovnog procesa	Otkrivanje poslovnog procesa	Unapređenje poslovnog procesa	Provera usaglašenosti	Analiza društvenih mreža	/
----------	--	------------------------------------	-------------------------------------	--------------------------	--------------------------------	---

Potrebno je napomenuti da smernice uključuju i tabelu 64, gde su pružene smernice za odabir algoritma i tehnike automatskog otkrivanja poslovnih procesa u zavisnosti od tipa automatskog otkrivanja poslovnih procesa koji se primenjuje, nastale na osnovu deskriptivne statistike posmatranih varijabli primarnih studija.

Tabela 64. Smernice za odabir algoritma i tehnike različitih tipova automatskog otkrivanja poslovnih procesa

Tip AOPP	Algoritmi i tehnike			
	Rang			
	1	2	3	4
Otkrivanje poslovnog procesa	Fazi algoritam	Induktivni algoritam	Heuristički algoritam	<i>Celonis</i> algoritam otkrivanja procesa
Unapredjenje poslovnog procesa	Fazi algoritam	Heuristički algoritam	Induktivni algoritam	<i>Celonis</i> tehnike unapređenja procesa
Provera usaglašenosti	<i>ProM</i> dodatak za proveru usaglašenosti	/	/	/
Analiza društvenih mreža	<i>ProM</i> dodatak za analizu društvenih mreža	/	/	/
Prediktivno AOPP	Bajezove mreže	Diskretna simulacija događaja	LSTM neuronske mreže	/

5.2. Diskusija rezultata istraživačkog modela II

Drugi deo doktorske disertacije je težio da odgovori na istraživačko pitanje: Koje tehnike čišćenja treba da se primene na dnevnik izvršenja događaja u zavisnosti od problema kvaliteta podataka?

Kako bi se ispitali problemi kvaliteta podataka koji se javljaju u dnevnicima izvršenja događaja, izvršen je pregled literature koji je izvršio sintezu svih problema kvaliteta i obrazaca imperfekcije definisanih u literaturi. Zatim je izvršen sistematski pregled literature tehnika čišćenja dnevnika izvršenja događaja koji se primenjuju kako bi se problemi kvaliteta podataka umanjili ili uklonili.

Na osnovu uvida u teorijske osnove, razvijen je istraživački model 2 koji sadrži jednu hipotezu o postojanju mogućnosti da se na osnovu određenog problema kvaliteta podataka odredi pogodna tehnika za rešavanje tog problema. Prikupljanje podataka za testiranje navedene hipoteze je sprovedeno putem upitnika, ispitivanjem istraživača i praktikanata iz oblasti automatskog otkrivanja poslovnih procesa. Rezultati analize podataka prikupljenih upitnikom se dalje mogu tumačiti kako bi se stekao značajniji uvid i definisale smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu.

5.2.1. Diskusija razvoja mernog instrumenta i demografije ispitanika

Upitnik je, kao merni instrument istraživačkog modela 2, definisan na osnovu posmatrane dve dimenzije: problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja. Na osnovu izvršenih pregleda literature, utvrđene su stavke upitnika koje određuju dimenzije istraživanja. Upitnik je sadržao demografska pitanja o ispitanicima, pitanja o značaju i učestalosti primene problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja, kao i pitanje o odabiru tehnike čišćenja u zavisnosti od problema kvaliteta podataka.

Prilikom deskriptivne statističke analize demografskih karakteristika ispitanika, doneti su značajni zaključci o stanju u oblasti automatskog otkrivanja poslovnih procesa. Pre svega, interesantno je zaključiti da se više od pola ispitanika bavi automatskim otkrivanjem poslovnih procesa u periodu 1 - 5 godina, što pokazuje da se tek u poslednjih 5 godina javlja povećano interesovanje i potreba na tržištu rada. Takođe, s obzirom da automatsko otkrivanje poslovnih procesa potiče sa Tehnološkog univerziteta u Ajndhovenu, kao oblast se prvo primenjivala na univerzitetima u Evropi. Stoga je bilo značajno ispitati da li, nakon 10 godina postojanja oblasti, postoji razlika u pogledu uloge koju ispitanici imaju i vremenskog periodom kojim se bave automatskim otkrivanjem poslovnih procesa. Rezultat Hi-kvadrat testa je pokazao da postoji statistički značajna razlika, gde većina istraživača ima značajno iskustvo u automatskom otkrivanju poslovnih procesa, koje traje više od 10 godina. S druge strane, većina praktičara ima iskustvo od 1-5 godina, što potvrđuje da je primena procesnog rudarstva u komercijalnim projektima još uvek nova i da se širi.

Veliki broj učesnika trenutno radi u Evropi (71%), kao što se i očekivalo, iz razloga što automatsko otkrivanje poslovnih procesa potiče sa Tehnološkog univerziteta u Ajndhovenu. Shodno tome, tamo je razvijen prvi istaknuti alat za automatsko otkrivanje poslovnih procesa. *ProM* je trenutno drugi najčešće korišćeni softverski alat za automatsko otkrivanje poslovnih procesa uopšte i treći najčešće korišćeni softverski alat za čišćenje dnevnika izvršenja događaja, što znači da i dalje ima značajan udeo na tržištu softverskih alata. Međutim, Nemačka trenutno zapošljava najviše praktičara i istraživača procesnog rudarstva, što je posledica komercijalnog uspeha *Celonis* alata za automatsko otkrivanje poslovnih procesa. *Celonis* ima sedište u Minhenu i često je korišćeni softver za automatsko otkrivanje poslovnih procesa i čišćenje dnevnika izvršenja događaja. Nemačka takođe ima visoko kvalitetan sistem obrazovanja i istraživanja, pri čemu se automatsko otkrivanje poslovnih procesa proučava na univerzitetima i u istraživačkim centrima, stvarajući veliki broj istraživača automatskog otkrivanja poslovnih procesa. S obzirom na softverske alate, *PM4Py* se može istaći kao alat koji se često koristi za pripremnu obradu podataka o događajima, kao i za otkrivanje i poboljšanje procesa.

Zanimanje ispitanika je zanimljiva tema, pošto je automatsko otkrivanje poslovnih procesa disciplina u nastajanju i poslednjih godina se primenjuje u industriji. Zanimljivo je primetiti da su postojeća zanimanja u domenu računarskih nauka, kao što su analitičar podataka, naučnik podataka i softverski inženjer, veoma tražena u oblasti automatskog otkrivanja poslovnih procesa. Veliki broj analitičara podataka i naučnika podataka potvrđuje da je kvalitet podataka prepoznat kao ključni aspekt projekata automatskog otkrivanja poslovnih procesa. Štaviše, analitičar automatskog otkrivanja poslovnih procesa je nova pozicija koja se nudi mladim praktičarima u organizacijama fokusiranim na upravljanje poslovnim procesima i konsalting, što pokazuje da se tržište automatskog otkrivanja poslovnih procesa širi i dobija na značaju.

Kako bi ispitale razlike između uloge ispitanika i odabira softverskog alata za primenu automatskog otkrivanja poslovnih procesa i softverskih alata za obradu podataka, primenjen je Hi-kvadrat test nezavisnosti, gde je zaključeno da istraživači, više nego druge grupe, koriste alate *ProM*, *Fluxicon Disco* i *Apromore*. Ovaj podatak je značaj iz razloga što se stiče uvid u to da istraživači koriste prve softverske alate iz oblasti, koji su razvijeni na univerzitetima. Ovi softverski alati sadrže same algoritme i tehnike koji su istraživači razvili i pružaju širok spektar različitih, vrlo specifičnih analiza. Sa druge strane, praktikanti dominantno koriste komercijalni alat *Celonis*, koji ima ugrađene algoritme i tehnike obrade podataka i analize poslovnih procesa i gde se ne zahteva od korisnika da razume logiku tehnologije koja se koristi za analizu.

5.2.2. Diskusija analize značaja i učestalosti problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja

Značaj i učestalost problema kvaliteta podataka se najbolje mogu diskutovati kroz rezultate IPA analize. IPA analiza o percipiranom značaju i učestalosti susreta sa problemima kvaliteta podataka dnevnika događaja je pre svega ukazala na najvažnije probleme kvaliteta podataka, odnosno probleme svrstane u kvadrant „Visok prioritet”. Očekivano, problemi koji se ovde javljaju su problem nedostajućih podataka o događajima, koji dovodi do pojave obrasca raštrkanog događaja, i problem nedostajućih podataka o vezi između događaja i slučaja procesa, koji dovodi do pojave obrasca neuhvatljivog slučaja. Oba problema su veoma visokog stepena značajnosti, jer obrazac raštrkanog događaja implicira da su psotoje događaji koji su izvršeni u procesu, ali su u dnevniku izvršenja događaja zabeleženi tako da nisu jasno uočljivi. Na osnovu takvog dnevnika izvršenja događaja se modeluje poslovni proces koji ne sadrži sve aktivnosti realnog sistema. Obrazac neuhvatljivog slučaja stvara još uticajnije probleme na rezultate analize, jer se zbog njegovog postojanja u dnevniku izvršenja događaja ne ispunjava osnovna pretpostava analize, a to je da svaki događaj pripada određenom slučaju procesa.

Takođe, problem kvaliteta visokog prioriteta su oni koji se tiču netačnih i nepreciznih vremenskih odrednica. Ovaj rezultat je takođe očekivan, jer vremenske odrednice direktno utiču na strukturu i tok događaja jednog modela procesa. Vremenske odrednice su čest problem kvaliteta podataka dnevnika izvršenja događaja, jer se često generišu na osnovu senzora i uređaja, koji mogu da budu u otkazu ili da su konfigurisani tako da ne beleže vreme izvršavanje aktivnosti niskog nivoa detaljnosti. Problemi sa nedostajućim, netačnim ili nepreciznim vremenskim odrednicama imaju značajne implikacije na analizu podataka i tumačenje rezultata automatskog otkrivanja poslovnih procesa. Shodno tome, postaje izazovno odrediti tačan redosled događaja, a svaka analiza učinka zasnovana na vremenu postaje nepouzdana. Učesnici su bili svesni ovih implikacija i shodno tome ocenili važnost ovog problema kvaliteta podataka. Takođe, verovatnoća greške je velika i u slučajevima kada se vreme izvršavanja aktivnosti unosi od strane korisnika informacionog sistema. Problem velikih količina podataka i kompleksnosti podataka, kao i različitog nivoa detaljnosti aktivnosti je takođe percipiran kao problem visokog prioriteta, jer informacioni sistemi generišu velike količine podataka o izvršavanju aktivnosti.

Učestalost problema kvaliteta podataka koji se susreću razlikuje se od percipirane važnosti, i zato je IPA matrica dala značajna zapažanja. Kvadrant „nizak prioritet“ sadrži probleme kvaliteta podataka koji se smatraju irelevantnim i koji se često susreću, na primer, neprecizni i netačni nazivi aktivnosti, neprecizni i netačni atributi slučaja/događaja, netačni podaci u vezi sa resursima i neprecizni podaci o odnosu između slučajeva i događaja. Zaključak je da ova pitanja kvaliteta podataka ne bi trebalo da budu u fokusu istraživanja ili poboljšanja i da je rad na njihovim rešenjima manje značajan od drugih pitanja kvaliteta podataka. Zanimljivo je primetiti da prkanti i istraživači smatraju da su nazivi i oznake naziva aktivnosti manje važni, iako zagađene i iskrivljeni nazivi aktivnosti mogu dovesti do modelovanja duplih događaja. Njihova kategorizacija u kvadrant „nizak prioritet“ je veća jer se ne susreću često, a njihov značaj je nešto ispod proseka.

Kvadrant „Tako nastavi“ sadrži probleme sa kvalitetom podataka koji se smatraju važnim, ali se ne susreću često. Pošto je poželjno svesti na minimum pojavu problema kvaliteta podataka, ovaj kvadrant sadrži pitanja koja bi trebalo da budu u umerenom fokusu daljih istraživanja. Pomenuti problemi sa kvalitetom podataka su nedsitajuća vremenska oznaka i netačani podaci o slučaju procesa, događaja i odnosa između njih. S druge strane, kvadrant „Srednji prioritet“ sadrži česta, ali manje važna pitanja. Pošto ih učesnici ne percipiraju kao važne, ne zahtevaju dodatni fokus. Pomenuti problemi kvaliteta podataka su irelevantni podaci o slučaju i događaju, nedostajući podaci o resursima i nedostajući atributi slučaja/događaja. Irelevantni podaci o resursima se smatraju nevažnim, jer nisu neophodni za sprovođenje tehnika automatskog otkrivanja poslovnih procesa.

IPA analiza je obavljena i za tehnike čišćenja dnevnika izvršenja događaja. Nekoliko kategorija tehnika čišćenja dnevnika izvršenja događaja, tj. tehnike popravke dnevnika podataka, tehnike mašinskog učenja, veštačke inteligencije i istraživanja i eksploatacije podataka i tehnike zasnovane na poravnanju, nalaze se u kvadrantu „nizak prioriteta“, što znači da se smatraju manje važnim i ne koriste se često. Razlog za mali značaj koji im ispitanici pridaju može biti u tome što su one specifične tehnike

obrade podataka, čija upotreba nije jednostavna i sa kojima nisu svi ispitanici upoznati. Kvadrant „srednji prioritet“ ne sadrži nikakve tehnike, što znači da ispitanici ne koriste tehnike koje ne smatraju važnim. Klasterovanje putanja, filtriranje putanja/događaja i apstrakcija događaja su tehnike koje se pravilno primenjuju, pošto se smatraju važnim i često se primenjuju. Kvadrant „visok prioritet“ treba da sadrži problematične primene tehnika čišćenja dnevnika izvršenja događaja, gde su neke tehnike važne, ali se ne primenjuju. Međutim, takve tehnike ne postoje, što znači da su istraživači i praktikanti svesni značaja određenih tehnika i da ih primenjuju u skladu sa tim.

U pogledu dodatnih analiza koje se tiču značaja i učestalosti problema kvaliteta podataka, analiza varijansi je pružila zanimljiva zapažanja. Pokazalo se da postoji razlika između ispitanika različitih uloga u pogledu značaja koji pridaju problemima kvaliteta podataka. Naime, praktikanti smatraju manje značajnim od ostalih uloga probleme netačnih podataka o vezi događaja i slučaja i sve vrste problema koji se tiču naziva aktivnosti. Navedeni problemi kategorisani kao visokog prioriteta od strane ukupnog broja ispitanika, te se zaključuje da praktikanti nemaju dovoljno znanja o značaju, manifestaciji i uticaju samih problema kvaliteta podataka na rezultate analize procesa. Ovo zapažanje se može povezati sa činjenicom da praktikanti u velikoj meri koriste softverski alat *Celonis*, koji ne zahteva ekstenzivno razumevanje same pripreme podataka za analizu i pozadinske logike tehnika automatskog otkrivanja poslovnih procesa. Takođe, može se napraviti veza i sa njihovim isustvom koje je uglavnom u kategoriji 1-5 godina. Dalje, u pogledu učestalosti susreta različitih uloga sa problemima kvaliteta podataka, analiza varijansi je pokazala da ispitanici koji pripadaju grupi istraživača i praktikanata mnogo češće susreću raznolike probleme kvaliteta podataka, nego samo istraživači ili samo praktikanti. Može se zaključiti da ispitanici koji pripadaju toj grupi imaju najraznovrsnije i najpotpunije znanje iz oblasti automatskog otkrivanja poslovnih procesa.

U pogledu dodatnih analiza koje se tiču značaja i učestalosti primene tehnika čišćenja dnevnika izvršenja događaja, analiza varijansi je pružila zanimljiva zapažanja. Pokazalo se da postoji razlika između ispitanika različitih uloga u pogledu značaja koji pridaju problemima kvaliteta podataka. Naime, istraživači manji značaj nekog ostale grupe pridaju ugrađenim tehnikama obrade podataka. Ove tehnike su ugrađene u algortime koji otkrivaju modele procesa i mogućnosti su im ograničene, u smislu da omogućavaju filtriranje aktivnosti samo na osnovu njihove frekvencije pojavljivanja i frekvencije njihovih veza sa drugim aktivnostima, čineći ih jednostavnim za korišćenje. Istraživači su ipak više naklonjeni kompleksnijim tehnikama čišćenja dnevnika izvršenja događaja nego praktikanti, koji ne pridaju značaj kompleksnijim tehnikama poput abstrakcije događaja i klasterovanja putanja. Što se tiče učestalosti primene tehnika izvršenja događaja, potvrđuju se prethodni zaključci, gde istraživači primenjuju tehnike popravke dnevnika izvršenja događaja više nego ostale grupe, dok praktikanti više primenjuju ugrađenu obradu podataka.

5.2.3. Diskusija rezultata ispitivanja hipoteze istraživačkog modela II

Istraživački model 2 sadrži dimenzije problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja i hipotezu koja pretpostavlja postojanje veze između tih dimenzija. U nastavku je dat komentar na rezultate ispitane hipoteze predstavljenih u **Error! Reference source not found..**

H6: Moguće je odrediti pogodnu tehniku čišćenja dnevnika izvršenja događaja u zavisnosti od problema kvaliteta podataka.

Rezultati Pirsonovog testa korelacije pokazali su da postoji statistički značajna veza između problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja, sa slabim i umerenim korelacionim koeficijentima.

Hi-kvadrat test je takođe potvrdio postojanje veze između posmatranih dimenzija. Na osnovu navedenog, hipoteza H6 je **prihvaćena**.

U pogledu istraživanja doktorske disertacije, prihvatanje hipoteze H6 dovodi do praktičnih implikacija istraživačkog modela 2, sagledane kroz kreirane smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu. Ove smernice pružaju menadžerima projekata i svim zainteresovanim stranama projekata automatskog istraživanja poslovnih procesa podlogu za donošenje

odluka o odabiru konkretne tehnike čišćenja dnevnika izvršenja događaja, na osnovu čega je moguće planirati tok izvršavanja projekta i alokaciju resursa.

5.2.4. Smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu

Smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu definisane su na osnovu snaga korelacije između problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja, kao i tehnike iz tabele kontingencije koje su primenjene u 15% ili više slučajeva radi rešavanja nekog od problema kvaliteta podataka. Dodatno, u smernice su uključeni samo problemi kvaliteta podataka koji od strane ispitanika procenjeni kao problemi od visokog značaja. Smernice su predstavljene u tabeli 65, gde za svaki problem kvaliteta podataka postoji utvrđen rang tehnike koja treba da se primeni, gde rang 1 predstavlja tehniku najvišeg prioriteta primene.

Tabela 65. Smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu

Problemi kvaliteta podataka	Kategorije tehnika čišćenja dnevnika izvršenja događaja		
	Rang		
	1	2	3
Nedostajući podaci: Slučaj	Klasterovanje putanja	Tehnike popravke dnevnika izvršenja događaja	Filtriranje putanje/događaja
Nedostajući podaci: Događaj (Raštrkani događaj)	Filtriranje putanje/događaja	Tehnike popravke dnevnika izvršenja događaja	Abstrakcija događaja
Nedostajući podaci: Veza (Neuhvatljiv slučaj)	Klasterovanje putanja	Abstrakcija događaja	Filtriranje putanje/događaja
Netačni podaci: Vremenska odrednica (Hvatanje događaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj)	Tehnike popravke dnevnika izvršenja događaja	Filtriranje putanje/događaja	/
Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj)	Tehnike popravke dnevnika izvršenja događaja	Filtriranje putanje/događaja	/
Količina podataka, nivo detaljnosti događaja, kompleksnost	Filtriranje putanje/događaja	Mašinsko učenje, veštačka inteligencija i istraživanje i eksploatacija podataka	Abstrakcija događaja

6. ZAKLJUČNA RAZMATRANJA I PRAVCI DALJIH ISTRAŽIVANJA

Doktorska disertacija je težila da umani probleme oblasti automatskog otkrivanja poslovnih procesa koji se tiču čišćenja dnevnika izvršenja događaja, benčmarkinga, unapređenja upotrebljivosti za početnike i unapređenja razumljivosti za početnike. Cilj doktorske disertacije je ispunjen, jer su kroz dva konkurentna istraživačka modela definisane smernice za planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa u zavisnosti od vrste industrije i cilja analize i smernice za planiranje i izvođenje pripreme dnevnika izvršenja događaja za analizu.

Teorijski doprinos doktorske disertacije se ogleda pre svega kroz rezultate sistematskih pregleda literature koji su sprovedeni radi utvrđivanja ključnih faktora istraživanja i postavljanja hipoteza, a koji se tiču primene automatskog otkrivanja poslovnih procesa u industriji i primene tehnika čišćenja dnevnika izvršenja događaja radi rešavanja različitih problema kvaliteta podataka.

Takođe, sami rezultati analize podataka kroz oba istraživačka modela daju značajni teorijski doprinos. Istraživački model 1 pruža sintezu podataka o primeni automatskog otkrivanja poslovnih procesa u 144 različite studije slučaja, sa sintezom podataka o industrijama, ciljevima analize, analiziranim poslovnim procesima, primenjenim tipovima automatskog otkrivanja poslovnih procesa, korišćenim softverskim alatima i konkretnim nazivima primenjenih algoritama i tehnika. Istraživački model 2 pruža sintezu podataka o demografskim karakteristikama istraživača i praktikanata automatskog otkrivanja poslovnih procesa i uvid u najznačajnije probleme kvaliteta podataka, tehnike čišćenja dnevnika izvršenja događaja i softverske alate obrade podataka.

Istraživanje sprovedeno kroz istraživački model 1 je odgovorilo na istraživačko pitanje IP1: Koji faktori projekata analize i automatskog otkrivanja poslovnih procesa (projektna karakteristika) utiču na odabir tipova, i algoritama/tehnika automatskog otkrivanja poslovnih procesa?

Rezultati su pokazali da su ključni faktori primene automatskog otkrivanja poslovnih procesa vrsta industrije, vrsta poslovnog procesa, cilj analize, tip automatskog otkrivanja poslovnih procesa, kao i algoritmi i tehnike iz oblasti. Statistički značajna povezanost je potvrđena između vrste industrije i vrste poslovnog procesa, vrste industrije i cilja analize, cilj analize i tipa automatskog otkrivanja poslovnih procesa, kao i cilj analize i vrste algoritama i tehnika. Na osnovu rezultata ispitivanja hipoteza istraživačkog modela 1, definisane su smernice za planiranje i izvođenje projekata automatskog otkrivanja poslovnih procesa u zavisnosti od vrste industrije i cilja analize, kao i smernice za odabir algoritama i tehnika u zavisnosti od tipa automatskog otkrivanja poslovnih procesa. Praktične implikacije su detaljno opisane u okviru diskusije rezultata istraživanja, ali najvažnija jeste korist koju praktikanti i istraživači imaju od definisanih smernica, kao alat za donošenje odluka prilikom planiranja izvođenja projekata automatskog otkrivanja poslovnih procesa i odabira pogodnih tipova analize u zavisnosti od scenarija u okviru kojeg se posmatrani proces odvija.

Istraživanje sprovedeno kroz istraživački model 2 je odgovorilo na istraživačko pitanje IP2: Koje tehnike čišćenja treba da se primene na dnevnik izvršenja događaja u zavisnosti od problema kvaliteta podataka?

Rezultati pokazuju uočenu važnost i učestalost susreta i upotrebe problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja, kao i IPA matricu odnosa između učestalosti i značaja, za obe dimenzije istraživanja. Nadalje, istraživanje pokazuje statistički značajnu vezu problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja. Pored toga, unakrsna analiza daje pregled ponašanja učesnika prilikom odabira tehnika čišćenja dnevnika izvršenja događaja u odnosu na problem kvaliteta podataka, nudeći značajne informacije za praktikante automatskog otkrivanja poslovnih procesa. Praktične implikacije su detaljno opisane u okviru diskusije rezultata istraživanja, ali najznačajnijim praktičnim doprinosom se mogu smatrati preporuke o izboru tehnika čišćenja dnevnika izvršenja događaja za probleme kvaliteta podataka visokog prioriteta, nudeći praktične smernice za praktikante i istraživače.

Istraživanje u okviru ove disertacije je prezentovano uzimajući u obzir određena ograničenja. Ukoliko se posmatraju nedostaci doktorske disertacije, treba napomenuti da se Pirsonov Hi-kvadrat test primenjuje radi ispitivanja hipoteza oba istraživačka modela zbog kategoričnosti prikupljenih podataka. Međutim, važno je napomenuti da iako statistički značajan rezultat ukazuje na vezu između varijabli, dalja analiza i interpretacija mogu biti potrebne da bi se razumela priroda i snaga tog odnosa. Pored toga, prema saznanjima autora, nije sprovedeno slično prethodno istraživanje, te nije moguće uporediti adekvatnost dobijenih rezultata. Štaviše, uzorak istraživačkog modela 2 se sastojao od istraživača i praktikanata svih nivoa iskustva u obradi podataka, osim onih koji su spadali u kategoriju „loše“. Uzorak korektno predstavlja zajednicu automatskog otkrivanja poslovnih procesa i njihove preporuke. Međutim, veći procenat „odličnog“ iskustva u poznavanju obrade podataka bi bio pogodniji.

Dalja istraživanja u pogledu istraživačkog modela 1 mogu biti fokusirana na dodatno ispitivanje hipoteze H3, koja je odbačena prilikom testiranja. Naime, hipoteza o vezi između cilja analize i vrste poslovnog procesa je odbačena. Mogući razlog za odbacivanje hipoteze može biti veliki

broj vrsta poslovnog procesa, te bi dalji rad na ovu temu mogao da grupiše vrste poslovnih procesa na višem nivou abstrakcije, po unapred određenom kriterijumu. Na taj način definisao bi se manji broj vrsta poslovnih procesa i povećala pouzdanost statističkih testova koji se primenjuju nad ovom varijablom. Ovakav pristup bi zahtevao i dodatnu analizu samih poslovnih procesa, njihovih ključnih osobina i sličnosti na osnovu kojih se mogu grupisati.

Dalja istraživanja u pogledu istraživačkog modela 2 mogu, s obzirom na potvrđene razlike u pogledu percepcije i učestalosti susreta i primene problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja od strane ispitanika različitih uloga, da podele uzorak na dva dela, u zavisnosti od uloge ispitanika. Zatim se može ponovo testirati hipoteza o povezanosti problema kvaliteta podataka i tehnika čišćenja dnevnika izvršenja događaja iz dva različita ugla, istraživača i praktikanata. Takođe, na osnovu tabele kontingencije definisane tokom unakrsne analize, moguće je kreirati smernice za planiranje obrade podataka iz ugla istraživača i ugla praktikanta i zatim te smernice uporediti sa smernicama definisanim u doktorskoj disertaciji.

7. LITERATURA

1. van der Aalst, W.M.P.; Adriansyah, A.; Alves De Medeiros, A.K.; Arcieri, F.; Baier, T.; Blickle, T.; Chandra Bose, J.; Van Den Brand, P.; Brandtjen, R.; Buijs, J.; et al. *Process Mining Manifesto*; 2012; pp. 169–194;.
2. van der Aalst, W.M.P. Process Mining: A 360 Degree Overview. In *Process Mining Handbook*; van der Aalst, W.M.P., Rubin, V., Eds.; Springer Cham, 2022; pp. 3–36.
3. van der Aalst, W. *Process Mining: Data Science in Action*; 2nd ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2016; ISBN 978-3-662-49850-7.
4. Dakic, D.; Stefanovic, D.; Cosic, I.; Lolic, T.; Medojevic, M. Business Process Mining Application: A Literature Review. In Proceedings of the Proceedings of the 29th DAAAM International Symposium; Katalinic, B., Ed.; DAAAM International: Vienna, Austria, 2018; pp. 0866–0875.
5. Reinkemeyer, L. Status and Future of Process Mining: From Process Discovery to Process Execution. In *Process Mining Handbook*; van der Aalst, W., Carmona, J., Eds.; Springer, 2022; Vol. 448, pp. 405–415.
6. Kerremans, M.; Srivastava, T.; Choudhary, F. *Gartner Market Guide for Process Mining*; 2021;
7. Bose, R.P.J.C.; Mans, R.S.; Van Der Aalst, W.M.P. Wanna Improve Process Mining Results? In Proceedings of the Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2013 - 2013 IEEE Symposium Series on Computational Intelligence, SSCI 2013; 2013; pp. 127–134.
8. Suriadi, S.; Andrews, R.; ter Hofstede, A.H.M.; Wynn, M.T. Event Log Imperfection Patterns for Process Mining: Towards a Systematic Approach to Cleaning Event Logs. *Inf Syst* **2017**, *64*, 132–150, doi:10.1016/j.is.2016.07.011.
9. Andrews, R.; Suriadi, S.; Ouyang, C.; Poppe, E. Towards Event Log Querying for Data Quality: Let's Start with Detecting Log Imperfections. In Proceedings of the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer Verlag, 2018; Vol. 11229 LNCS, pp. 116–134.
10. Fischer, D.A.; Goel, K.; Andrews, R.; van Dun, C.G.J.; Wynn, M.T.; Röglinger, M. Towards Interactive Event Log Forensics: Detecting and Quantifying Timestamp Imperfections. *Inf Syst* **2022**, *109*, doi:10.1016/j.is.2022.102039.
11. Verhulst, R. Evaluating Quality of Event Data within Event Logs: An Extensible Framework, Eindhoven University of Technology: Eindhoven, 2016.
12. Vugs, L.; van Asseldonk, M.; van Son, N. Lumigi: Shining Light on Your Process Data. In Proceedings of the 3rd International Conference on Process Mining (ICPM 2021); 2021.
13. Khannat, A.; Sbai, H.; Kjiri, L. Event Logs Pre-Processing for Configurable Process Discovery: Ontology-Based Approach. In Proceedings of the Colloquium in Information Science and Technology, CIST; Institute of Electrical and Electronics Engineers Inc., June 5 2020; Vol. 2020-June, pp. 139–144.
14. Marin-Castro, H.M.; Tello-Leal, E. Event Log Preprocessing for Process Mining: A Review. *Applied Sciences (Switzerland)* **2021**, *11*.
15. Aguirre, S.; Parra, C.; Sepúlveda, M. Methodological Proposal for Process Mining Projects. *International Journal of Business Process Integration and Management* **2017**, *8*, 102, doi:10.1504/IJBPM.2017.083793.
16. van Eck, M.L.; Lu, X.; Leemans, S.J.J.; van der Aalst, W.M.P. PM²: A Process Mining Project Methodology. In *Lecture Notes in Computer Science*; Zdravkovic, J., Kirikova, M., Johannesson, P., Eds.; Springer, Berlin: Stockholm, Sweden, 2015; Vol. 9097, pp. 297–313.

17. van der Aalst, W. Process Mining: Discovering and Improving Spaghetti and Lasagna Processes. In Proceedings of the 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM); IEEE: Paris, France, April 2011; pp. 1–7.
18. Van der Heijden, T. Process Mining Project Methodology: Developing a General Approach to Apply Process Mining in Practice. Master's thesis, School of Industrial Engineering, TUE, Netherlands: Eindhoven, Netherlands, 2012.
19. Bozkaya, M.; Gabriels, J.; van der Werf, J.M. Process Diagnostics: A Method Based on Process Mining. In Proceedings of the 2009 International Conference on Information, Process, and Knowledge Management; IEEE, February 2009; pp. 22–27.
20. Mărușter, L.; van Beest, N.R.T.P. Redesigning Business Processes: A Methodology Based on Simulation and Process Mining Techniques. *Knowl Inf Syst* **2009**, *21*, 267–297, doi:10.1007/s10115-009-0224-0.
21. Cho, M.; Song, M.; Yoo, S. A Systematic Methodology for Outpatient Process Analysis Based on Process Mining. **2015**, *22*, 480–493, doi:10.1007/978-3-319-08222-6_3.
22. Kim, D. Development of Practical Guidelines for Healthcare Process Mining. *ICIC Express Letters, Part B: Applications* **2018**, *9*, 977–982.
23. Augusto, V.; Xie, X.; Prodel, M.; Jouaneton, B.; Lamarsalle, L. Evaluation of Discovered Clinical Pathways Using Process Mining and Joint Agent-Based Discrete-Event Simulation. In Proceedings of the 2016 Winter Simulation Conference (WSC); IEEE: Washington, DC, USA, December 2016; pp. 2135–2146.
24. Rojas, E.; Sepúlveda, M.; Munoz-Gama, J.; Capurro, D.; Traver, V.; Fernandez-Llatas, C. Question-Driven Methodology for Analyzing Emergency Room Processes Using Process Mining. *Applied Sciences* **2017**, *7*, 302, doi:10.3390/app7030302.
25. Rebuge, Á.; Ferreira, D.R. Business Process Analysis in Healthcare Environments: A Methodology Based on Process Mining. *Inf Syst* **2012**, *37*, 99–116, doi:10.1016/j.is.2011.01.003.
26. Dakic, D.; Stefanovic, D.; Lolic, T.; Narandzic, D.; Simeunovic, N. Event Log Extraction for the Purpose of Process Mining: A Systematic Literature Review. In *Springer Proceedings in Business and Economics*; Prostean, G., Lavios Villahoz, J., Brancu, L., Bakacsi, G., Eds.; Springer, Cham, 2020; pp. 299–312.
27. Stefanovic, D.; Dakic, D.; Stevanov, B.; Lolic, T. Process Mining in Manufacturing: Goals, Techniques and Applications. In *IFIP Advances in Information and Communication Technology*; Lalic, B., Majstorovic, V., Marjanovic, U., von Cieminski, G., Romero, D., Eds.; Springer, Cham, 2020; Vol. 591, pp. 54–62.
28. Stefanovic, D.; Dakic, D.; Stevanov, B.; Lolic, T.; Marjanovic, U. Process Mining in the Manufacturing Context: Review and Recommendations. *International Journal of Industrial Engineering: Theory, Applications and Practice* **2021**, *28*.
29. Dakic, D.; Stefanovic, D.; Vuckovic, T.; Zizakov, M.; Stevanov, B. Event Log Data Quality Issues and Solutions. *Mathematics* **2023**, *11*, 2858, doi:10.3390/math11132858.
30. Pallant, J. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*; 7th ed.; Routledge: London, 2011;
31. Martilla, J.; James, J. Importance-Performance Analysis. *The Journal of Marketing* **1977**, *41*, 77–79.
32. Pereira, G.B.; Santos, E.A.P.; Maceno, M.M.C. Process Mining Project Methodology in Healthcare: A Case Study in a Tertiary Hospital. *Network Modeling Analysis in Health Informatics and Bioinformatics* **2020**, *9*, doi:10.1007/s13721-020-00227-w.
33. Levy, D. Production Analysis with Process Mining Technology.

34. Greyling, B.T.; Jooste, W. The Application of Business Process Mining to Improving a Physical Asset Management Process: A Case Study. *South African Journal of Industrial Engineering* **2017**, *28*, doi:10.7166/28-2-1691.
35. Augusto, A.; Carmona, J.; Verbeek, E. Advanced Process Discovery Techniques. In *Process Mining Handbook*; van der Aalst, W., Carmona, J., Eds.; Springer, Cham, 2022; Vol. 448, pp. 76–108.
36. Carmona, J.; van Dongen, B.; Weidlich, M. Conformance Checking: Foundations, Milestones and Challenges. In *Process Mining Handbook*; van der Aalst, W., Carmona, J., Eds.; Springer, Cham, 2022; Vol. 448.
37. Mahendrawathi, E.R.; Zayin, S.O.; Pamungkas, F.J. ERP Post Implementation Review with Process Mining: A Case of Procurement Process. *Procedia Comput Sci* **2017**, *124*, 216–223, doi:10.1016/j.procs.2017.12.149.
38. Epure, E. V.; Espen Ingvaldsen, J.; Deneckere, R.; Salinesi, C. Process Mining for Recommender Strategies Support in News Media. In Proceedings of the 2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS); IEEE, June 2016; pp. 1–12.
39. Stefanini, A.; Aloini, D.; Dulmin, R.; Mininno, V. Linking Diagnostic-Related Groups (DRGs) to Their Processes by Process Mining. In Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies; 2016.
40. Vázquez-Barreiros, B., C.D., M.M., L.M., & B.D. Process Mining in IT Service Management: A Case Study. In Proceedings of the CEUR Workshop Proceedings; 2016; pp. 16–30.
41. Riz, G.; Santos, E.A.P.; Loures, E.D.F.R. Process Mining to Knowledge Discovery in Healthcare Processes. In *Proceedings of Advances in Transdisciplinary Engineering*; 2016; Vol. 4, pp. 1019–1028.
42. Aisa, V.; Kurniati, A.P.; Yanuar Firdaus, A.W. Evaluation of the Online Assessment Test Using Process Mining (Case Study: Intensive English Center). In Proceedings of the 2015 3rd International Conference on Information and Communication Technology (ICoICT); IEEE, May 2015; pp. 472–477.
43. Mahendrawathi, E.R.; Astuti, H.M.; Nastiti, A. Analysis of Customer Fulfilment with Process Mining: A Case Study in a Telecommunication Company. *Procedia Comput Sci* **2015**, *72*, 588–596, doi:10.1016/j.procs.2015.12.167.
44. Heber, E.; Hagen, H.; Schmollinger, M. Application of Process Mining for Improving Adaptivity in Case Management Systems. In *Lecture Notes in Informatics (LNI)*; 2015; Vol. 244, pp. 221–231.
45. Caetano, A.; Pinto, P.; Mendes, C.; da Silva, M.M.; Borbinha, J. Analysis of Business Processes with Enterprise Ontology and Process Mining. In *Lecture Notes in Business Information Processing*; Aveiro, D., Pergl, R., Valenta, M., Eds.; Springer: Cham, 2015; Vol. 2011, pp. 82–95.
46. Sztyler, T.; Völker, J.; Carmona, J.; Meier, O.; Stuckenschmidt, H. Discovery of Personal Processes from Labeled Sensor Data-An Application of Process Mining to Personalized Health Care. In Proceedings of the Proceedings of the International Workshop on Algorithms & Theories for the Analysis of Event Data: Brussels, Belgium, June 22-23, 2015; CEUR.WS.org: Brussels, Belgium, 2015; pp. 31–46.
47. Antonelli, D.; Bruno, G. Application of Process Mining and Semantic Structuring Towards a Lean Healthcare Network. In *IFIP Advances in Information and Communication Technology*; Springer: Cham, 2015; pp. 497–508.
48. van der Aalst, W.M.P.; Guo, S.; Gorissen, P. Comparative Process Mining in Education: An Approach Based on Process Cubes. In *Lecture Notes in Business Information Processing*; Springer: Berlin, 2015; Vol. 203, pp. 110–134.

49. Rubin, V.A.; Mitsyuk, A.A.; Lomazova, I.A.; van der Aalst, W.M.P. Process Mining Can Be Applied to Software Too! In Proceedings of the Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement; ACM: New York, NY, USA, September 18 2014; pp. 1–8.
50. Bautista, A.D.; Wangikar, L.; Akbar, S.M.K. Process Mining-Driven Optimization of a Consumer Loan Approvals Process. In *Lecture Notes in Business Information Processing*; La Rosa, M., Soffer, P., Eds.; Springer: Berlin, 2013; Vol. 132, pp. 219–220.
51. Kim, E.; Kim, S.; Song, M.; Kim, S.; Yoo, D.; Hwang, H.; Yoo, S. Discovery of Outpatient Care Process of a Tertiary University Hospital Using Process Mining. *Healthc Inform Res* **2013**, *19*, 42, doi:10.4258/hir.2013.19.1.42.
52. Rozinat, A.; de Jong, I.S.M.; Gunther, C.W.; van der Aalst, W.M.P. Process Mining Applied to the Test Process of Wafer Scanners in ASML. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **2009**, *39*, 474–479, doi:10.1109/TSMCC.2009.2014169.
53. Perimal-Lewis, L.; Teubner, D.; Hakendorf, P.; Horwood, C. Application of Process Mining to Assess the Data Quality of Routinely Collected Time-Based Performance Data Sourced from Electronic Health Records by Validating Process Conformance. *Health Informatics J* **2016**, *22*, 1017–1029, doi:10.1177/1460458215604348.
54. Park, J.; Jung, J.-Y.; Jung, W. The Use of a Process Mining Technique to Characterize the Work Process of Main Control Room Crews: A Feasibility Study. *Reliab Eng Syst Saf* **2016**, *154*, 31–41, doi:10.1016/j.res.2016.05.004.
55. Kurniati, A.P.; Atastina, I. Implementing Process Mining to Improve COBIT 5 Assessment Program or Managing Operations (Case Study: A University Blog). *J Theor Appl Inf Technol* **2015**, *72*, 191–198.
56. Huang, C.; Cai, H.; Li, Y.; Du, J.; Bu, F.; Jiang, L. A Process Mining Based Service Composition Approach for Mobile Information Systems. *Mobile Information Systems* **2017**, *2017*, 1–13, doi:10.1155/2017/3254908.
57. Alvarez, C.; Rojas, E.; Arias, M.; Munoz-Gama, J.; Sepúlveda, M.; Herskovic, V.; Capurro, D. Discovering Role Interaction Models in the Emergency Room Using Process Mining. *J Biomed Inform* **2018**, *78*, 60–77, doi:10.1016/j.jbi.2017.12.015.
58. ER, M.; Arsad, N.; Astuti, H.M.; Kusumawardani, R.P.; Utami, R.A. Analysis of Production Planning in a Global Manufacturing Company with Process Mining. *Journal of Enterprise Information Management* **2018**, *31*, 317–337, doi:10.1108/JEIM-01-2017-0003.
59. Greyling, B.T.; Jooste, W. THE APPLICATION OF BUSINESS PROCESS MINING TO IMPROVING A PHYSICAL ASSET MANAGEMENT PROCESS: A CASE STUDY. *South African Journal of Industrial Engineering* **2017**, *28*, doi:10.7166/28-2-1691.
60. Pérez-Castillo, R.; Weber, B.; de Guzmán, I.G.-R.; Piattini, M. Process Mining through Dynamic Analysis for Modernising Legacy Systems. *IET Software* **2011**, *5*, 304, doi:10.1049/iet-sen.2010.0103.
61. Samalikova, J.; Kusters, R.J.; Trienekens, J.J.M.; Weijters, A.J.M.M. Process Mining Support for Capability Maturity Model Integration-Based Software Process Assessment, in Principle and in Practice. *Journal of Software: Evolution and Process* **2014**, *26*, 714–728, doi:10.1002/smr.1645.
62. Cho, M.; Song, M.; Yoo, S. A Systematic Methodology for Outpatient Process Analysis Based on Process Mining. In; 2014; pp. 31–42.
63. Sedrakyan, G.; De Weerd, J.; Snoeck, M. Process-Mining Enabled Feedback: “Tell Me What I Did Wrong” vs. “Tell Me How to Do It Right.” *Comput Human Behav* **2016**, *57*, 352–376, doi:10.1016/j.chb.2015.12.040.

64. Sahlabadi DETECTING ABNORMAL BEHAVIOR IN SOCIAL NETWORK WEBSITES BY USING A PROCESS MINING TECHNIQUE. *Journal of Computer Science* **2014**, *10*, 393–402, doi:10.3844/jcssp.2014.393.402.
65. Jans, M.; van der Werf, J.M.; Lybaert, N.; Vanhoof, K. A Business Process Mining Application for Internal Transaction Fraud Mitigation. *Expert Syst Appl* **2011**, *38*, 13351–13359, doi:10.1016/j.eswa.2011.04.159.
66. Mărușter, L.; van Beest, N.R.T.P. Redesigning Business Processes: A Methodology Based on Simulation and Process Mining Techniques. *Knowl Inf Syst* **2009**, *21*, 267–297, doi:10.1007/s10115-009-0224-0.
67. De Weerd, J.; Schupp, A.; Vanderloock, A.; Baesens, B. Process Mining for the Multi-Faceted Analysis of Business Processes—A Case Study in a Financial Services Organization. *Comput Ind* **2013**, *64*, 57–67, doi:10.1016/j.compind.2012.09.010.
68. Partington, A.; Wynn, M.; Suriadi, S.; Ouyang, C.; Karnon, J. Process Mining for Clinical Processes. *ACM Trans Manag Inf Syst* **2015**, *5*, 1–18, doi:10.1145/2629446.
69. Rojas, E.; Sepúlveda, M.; Munoz-Gama, J.; Capurro, D.; Traver, V.; Fernandez-Llatas, C. Question-Driven Methodology for Analyzing Emergency Room Processes Using Process Mining. *Applied Sciences* **2017**, *7*, 302, doi:10.3390/app7030302.
70. Saraeian, S.; Shirazi, B. Process Mining-Based Anomaly Detection of Additive Manufacturing Process Activities Using a Game Theory Modeling Approach. *Comput Ind Eng* **2020**, *146*, 106584, doi:10.1016/j.cie.2020.106584.
71. Stertz, F.; Rinderle-Ma, S. Detecting and Identifying Data Drifts in Process Event Streams Based on Process Histories. In *Lecture Notes in Business Information Processing*; Cappiello, C., Ruiz, M., Eds.; Springer, Cham, 2019; Vol. 350, pp. 240–252.
72. Nagy, Z.; Werner-Stark, A.; Dulai, T. Using Process Mining in Real-Time to Reduce the Number of Faulty Products. In *Advances in Databases and Information Systems*; 2019; pp. 89–104.
73. Wang, Y.; Hulstijn, J.; Tan, Y. Towards Smart Manufacturing: Compliance Monitoring for Computational Auditing. *Research papers* **2018**, *93*.
74. Satitcharoenmuang, C.; Porouhan, P.; Nammakhunt, A.; Saguansakiyotin, N.; Premchaiswadi, W. Benchmarking Efficiency of Children's Garment Production Process Using Alpha and ILP Replayer Techniques. In *Proceedings of the 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*; IEEE, November 2017; pp. 1–7.
75. Bettacchi, A.; Polzonetti, A.; Re, B. Understanding Production Chain Business Process Using Process Mining: A Case Study in the Manufacturing Scenario. In *Lecture Notes in Business Information Processing*; Krogstie, J., Mouratidis, H., Su, J., Eds.; Springer, Cham, 2016; Vol. 249, pp. 193–203.
76. Yahya, B.N. The Development of Manufacturing Process Analysis: Lesson Learned from Process Mining. *Jurnal Teknik Industri* **2014**, *16*, doi:10.9744/jti.16.2.95-106.
77. DONG, C. Resource Modeling of Manufacturing Process and Critical Nodes Recognition Based on the Integration of Process Mining and Complex Network. *Journal of Mechanical Engineering* **2019**, *55*, 169, doi:10.3901/JME.2019.03.169.
78. Park, M.; Song, M.; Baek, T.H.; Son, S.; Ha, S.J.; Cho, S.W. Workload and Delay Analysis in Manufacturing Process Using Process Mining. In *Lecture Notes in Business Information Processing*; Bae, J., Suriadi, S., Wen, L., Eds.; Springer, Cham, 2015; Vol. 219, pp. 138–151.
79. Kong, T.; Seong, K.; Song, K.; Lee, K. Two-Mode Modularity Clustering of Parts and Activities for Cell Formation Problems. *Comput Oper Res* **2018**, *100*, 77–88, doi:10.1016/j.cor.2018.06.018.

80. Song, M.; Jang, Y.; Son, S.; Yahya, B.N.; Choi, S.; Hyeon, J.; Lee, B.; Sung, N. Process Mining for Manufacturing Process Analysis: A Case Study. In Proceedings of the Asia Pacific Conference on Business Process Management; Brisbane, Australia, 2014.
81. van Eck, M.L.; Sidorova, N.; van der Aalst, W.M.P. Enabling Process Mining on Sensor Data from Smart Products. In Proceedings of the 2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS); IEEE: Grenoble, France, June 2016; pp. 1–12.
82. Viale, P.; Benayadi, N.; Le Goc, M.; Pinaton, J. Modeling Large Scale Manufacturing Process from Timed Data. In Proceedings of the Proceedings of the 12th International Conference on Enterprise Information Systems; SciTePress - Science and Technology Publications, 2010; pp. 129–138.
83. Engel, R.; Krathu, W.; Zapletal, M.; Pichler, C.; Bose, R.P.J.C.; van der Aalst, W.; Werthner, H.; Huemer, C. Analyzing Inter-Organizational Business Processes. *Information Systems and e-Business Management* **2016**, *14*, 577–612, doi:10.1007/s10257-015-0295-2.
84. Mušič, G.; Rojec, P. Process Mining of Production Management Data for Improvement of Production Planning and Manufacturing Execution. In Proceedings of the European Modeling and Simulation Symposium; 2012.
85. Yang, H.; Park, M.; Cho, M.; Song, M.; Kim, S. A System Architecture for Manufacturing Process Analysis Based on Big Data and Process Mining Techniques. In Proceedings of the 2014 IEEE International Conference on Big Data (Big Data); IEEE, October 2014; pp. 1024–1029.
86. Diba, K.; Remy, S.; Pufahl, L. *Compliance and Performance Analysis of Procurement Processes Using Process Mining*;
87. Dakic, D.; Sladojevic, S.; Lolic, T.; Stefanovic, D. Process Mining Possibilities and Challenges: A Case Study. In Proceedings of the 2019 IEEE 17th International Symposium on Intelligent Systems and Informatics (SISY); IEEE, September 2019; pp. 000161–000166.
88. Rbigui, H.; Cho, C. Purchasing Process Analysis with Process Mining of a Heavy Manufacturing Industry. In Proceedings of the 2018 International Conference on Information and Communication Technology Convergence (ICTC); IEEE: Jeju, Korea, October 2018; pp. 495–498.
89. Knoll, D.; Reinhart, G.; Prüglmeier, M. Enabling Value Stream Mapping for Internal Logistics Using Multidimensional Process Mining. *Expert Syst Appl* **2019**, *124*, 130–142, doi:10.1016/j.eswa.2019.01.026.
90. Dišek, M.; Šperka, R.; Kolesár, J. Conversion of Real Data from Production Process of Automotive Company for Process Mining Analysis. In *Agent and Multi-Agent Systems: Technology and Applications*; Jezic, G., Kusek, M., Chen-Burger, YH., Howlett, R., Jain, L., Eds.; Springer, Cham, 2018; Vol. 74, pp. 223–233.
91. Schuh, G.; Gützlaff, A.; Schmitz, S.; van der Aalst, W.M.P. Data-Based Description of Process Performance in End-to-End Order Processing. *CIRP Annals* **2020**, *69*, 381–384, doi:10.1016/j.cirp.2020.03.013.
92. Park, J.; Lee, D.; Zhu, J. An Integrated Approach for Ship Block Manufacturing Process Performance Evaluation: Case from a Korean Shipbuilding Company. *Int J Prod Econ* **2014**, *156*, 214–222, doi:10.1016/j.ijpe.2014.06.012.
93. Meincheim, A.; Garcia, C. dos S.; Nievola, J.C.; Scalabrin, E.E. Combining Process Mining with Trace Clustering: Manufacturing Shop Floor Process - An Applied Case. In Proceedings of the 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI); IEEE, November 2017; pp. 498–505.

94. Hong Tu, T.B.; Song, M. Analysis and Prediction Cost of Manufacturing Process Based on Process Mining. In Proceedings of the 2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA); IEEE: Jeju, Korea, May 2016; pp. 1–5.
95. Ruschel, E.; Santos, E.A.P.; Loures, E. de F.R. Establishment of Maintenance Inspection Intervals: An Application of Process Mining Techniques in Manufacturing. *J Intell Manuf* **2020**, *31*, 53–72, doi:10.1007/s10845-018-1434-7.
96. Bhogal, R.; Garg, A. Anomaly Detection and Fault Prediction of Breakdown to Repair Process Using Mining Techniques. In Proceedings of the 2020 International Conference on Intelligent Engineering and Management (ICIEM); IEEE, June 2020; pp. 240–245.
97. Sánchez, M.; Exposito, E.; Aguilar, J. Implementing Self-* Autonomic Properties in Self-Coordinated Manufacturing Processes for the Industry 4.0 Context. *Comput Ind* **2020**, *121*, 103247, doi:10.1016/j.compind.2020.103247.
98. Choueiri, A.C.; Sato, D.M.V.; Scalabrin, E.E.; Santos, E.A.P. An Extended Model for Remaining Time Prediction in Manufacturing Systems Using Process Mining. *J Manuf Syst* **2020**, *56*, 188–201, doi:10.1016/j.jmsy.2020.06.003.
99. Fleig, C.; Augenstein, D.; Maedche, A. Process Mining for Business Process Standardization in ERP Implementation Projects – An SAP S/4 HANA Case Study from Manufacturing.; Nepal, S., Dumas, M., Pentland, B., Kumar, A., Mendling, J., de Leoni, M., van der Aalst, W., Conforti, R., Casati, F., Weber, B., Eds.; CEUR-WS, 2018; Vol. 2196, pp. 149–155.
100. Farooqui, A.; Bengtsson, K.; Falkman, P.; Fabian, M. From Factory Floor to Process Models: A Data Gathering Approach to Generate, Transform, and Visualize Manufacturing Processes. *CIRP J Manuf Sci Technol* **2019**, *24*, 6–16, doi:10.1016/j.cirpj.2018.12.002.
101. Albertetti, F.; Ghorbel, H. Workload Prediction of Business Processes -- An Approach Based on Process Mining and Recurrent Neural Networks. *Computer Science* **2020**.
102. Garcia, C. dos S.; Meinheim, A.; Filho, F.C.G.; Santos, E.A.P.; Scalabrin, E.E. Getting Insights to Improve Business Processes with Agility: A Case Study Using Process Mining. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC); IEEE: Bari, Italy, October 2019; pp. 1336–1343.
103. Neshastegaran, A.; Norouzifar, A.; Izadi, I. A Framework for Plant Topology Extraction Using Process Mining and Alarm Data.; Institute of Electrical and Electronics Engineers Inc., 2022; pp. 485–491.
104. Schuh, G.; Gützlaff, A.; Cremer, S.; Schopen, M. Understanding Process Mining for Data-Driven Optimization of Order Processing.; Ramsauer, C., Wolf, M., Hulla, M., Eds.; Elsevier B.V., 2020; Vol. 45, pp. 417–422.
105. Meinheim, A.; Garcia, C.D.S.; Nievola, J.C.; Scalabrin, E.E. Combining Process Mining with Trace Clustering: Manufacturing Shop Floor Process-an Applied Case.; IEEE Computer Society, 2018; Vol. 2017-November, pp. 498–505.
106. Rogge-Solti, A.; Mans, R.S.; van der Aalst, W.M.P.; Weske, M. Improving Documentation by Repairing Event Logs. *Lecture Notes in Business Information Processing* **2013**, *165 LNBIP*, 129–144, doi:10.1007/978-3-642-41641-5_10.
107. Rogge-Solti, A.; Mans, R.S.; Van Der Aalst, W.M.P.; Weske, M. Repairing Event Logs Using Timed Process Models. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2013**, *8186 LNCS*, 705–708, doi:10.1007/978-3-642-41033-8_89.
108. Shahzadi, S.; Fang, X.; Shahzad, U.; Ahmad, I.; Benedict, T. Repairing Event Logs to Enhance the Performance of a Process Mining Model. *Math Probl Eng* **2022**, *2022*, doi:10.1155/2022/4741232.

109. Lu, Y.; Chen, Q.; Poon, S.K. A Deep Learning Approach for Repairing Missing Activity Labels in Event Logs for Process Mining. *Information (Switzerland)* **2022**, *13*, doi:10.3390/info13050234.
110. Sim, S.; Bae, H.; Choi, Y.; (WINTEC), C. et al.; E.L.P.D.W.I. of T. Likelihood-Based Multiple Imputation by Event Chain Methodology for Repair of Imperfect Event Logs with Missing Data. In Proceedings of the 1st International Conference on Process Mining, ICPM 2019; Institute of Electrical and Electronics Engineers Inc., 2019; pp. 9–16.
111. Liu, Y.; Yang, L.; Ghasemkhani, A.; Livani, H.; Centeno, V.A.; Chen, P.-Y.; Zhang, J. Robust Event Classification Using Imperfect Real-World PMU Data. *IEEE Internet Things J* **2023**, *10*, 7429–7438, doi:10.1109/JIOT.2022.3177686.
112. Horita, H.; Kurihashi, Y.; Miyamori, N. Extraction of Missing Tendency Using Decision Tree Learning in Business Process Event Log. *Data (Basel)* **2020**, *5*, 1–12, doi:10.3390/data5030082.
113. Ramos-Gutiérrez, B.; Varela-Vaca, Á.J.; Ortega, F.J.; Gómez-López, M.T.; Wynn, M.T. *A NLP-Oriented Methodology to Enhance Event Log Quality*; Augusto, A., Gill, A., Nurcan, S., Reinhartz-Berger, I., Schmidt, R., Zdravkovic, J., Eds.; Online Conference, 2021; Vol. 421;.
114. Chen, Q.; Lu, Y.; Tam, C.S.; Poon, S.K. A Multi-View Framework to Detect Redundant Activity Labels for More Representative Event Logs in Process Mining. *Future Internet* **2022**, *14*, doi:10.3390/fi14060181.
115. Nguyen, P.; Slominski, A.; Muthusamy, V.; Ishakian, V.; Nahrstedt, K. Process Trace Clustering: A Heterogeneous Information Network Approach. In Proceedings of the 16th SIAM International Conference on Data Mining 2016, SDM 2016; Venkatasubramanian, S.C., Meira, W., Eds.; Society for Industrial and Applied Mathematics Publications, 2016; pp. 279–287.
116. Boltenhagen, M.; Chatain, T.; Carmona, J. Generalized Alignment-Based Trace Clustering of Process Behavior. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Aachen, 2019; Vol. 11522, pp. 237–257.
117. Lu, X.; Fahland, D.; Van Der Aalst, W.M.P. Interactively Exploring Logs and Mining Models with Clustering, Filtering, and Relabeling. In Proceedings of the CEUR Workshop Proceedings; Azevedo L. Cabanillas C., A.L., Ed.; CEUR-WS, 2016; Vol. 1789, pp. 44–49.
118. Liu, J.; Xu, J.; Zhang, R.; Reiff-Marganiec, S. A Repairing Missing Activities Approach with Succession Relation for Event Logs. *Knowl Inf Syst* **2021**, *63*, 477–495, doi:10.1007/s10115-020-01524-6.
119. Ceravolo, P.; Damiani, E.; Torabi, M.; Barbon, S. Toward a New Generation of Log Pre-Processing Methods for Process Mining. In *15th International Conference on Business Process Management, BPM 2017*; Barcelona, 2017; pp. 55–70.
120. Sadeghianasl, S. The Quality Guardian: Improving Activity Label Quality in Event Logs Through Gamification. In Proceedings of the 2022 Best Dissertation Award, Doctoral Consortium, and Demonstration and Resources Track at BPM, BPM-D 2022; Janiesch, C., C, D.F., Grisold, T., Kumar, A., Mendling, J., Pentland, B., Reijers, H., Winter, R., Weske, M., Eds.; CEUR-WS, 2022; Vol. 3216, pp. 1–5.
121. Tax, N.; Sidorova, N.; van der Aalst, W.M.P. Discovering More Precise Process Models from Event Logs by Filtering out Chaotic Activities. *J Intell Inf Syst* **2019**, *52*, 107–139, doi:10.1007/s10844-018-0507-6.
122. Conforti, R.; La Rosa, M.; Ter Hofstede, A.H.M. Filtering Out Infrequent Behavior from Business Process Event Logs. *IEEE Trans Knowl Data Eng* **2017**, *29*, 300–314, doi:10.1109/TKDE.2016.2614680.

123. van Zelst, S.J.; Fani Sani, M.; Ostovar, A.; Conforti, R.; La Rosa, M. Filtering Spurious Events from Event Streams of Business Processes. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Tallinn, 2018; Vol. 10816, pp. 35–52.
124. Huang, R.; Wang, J.; Song, S.; Lin, X.; Zhu, X.; Pei, J. Efficiently Cleaning Structured Event Logs: A Graph Repair Approach. *ACM Transactions on Database Systems* **2023**, *48*, doi:10.1145/3571281.
125. Wang, J.; Song, S.; Zhu, X.; Lin, X.; Sun, J. Efficient Recovery of Missing Events. *IEEE Trans Knowl Data Eng* **2016**, *28*, 2943–2957, doi:10.1109/TKDE.2016.2594785.
126. Song, S.; Huang, R.; Cao, Y.; Wang, J. Cleaning Timestamps with Temporal Constraints. *VLDB Journal* **2021**, *30*, 425–446, doi:10.1007/s00778-020-00641-6.
127. Sani, M.F.; van Zelst, S.J.; van der Aalst, W.M.P. Improving Process Discovery Results by Filtering Outliers Using Conditional Behavioural Probabilities. *Lecture Notes in Business Information Processing* **2018**, *308*, 216–229, doi:10.1007/978-3-319-74030-0_16.
128. Fani Sani, M.; van Zelst, S.J.; van der Aalst, W.M.P. Repairing Outlier Behaviour in Event Logs. *Lecture Notes in Business Information Processing* **2018**, *320*, 115–131, doi:10.1007/978-3-319-93931-5_9.
129. Song, W.; Xia, X.; Jacobsen, H.-A.; Zhang, P.; Hu, H.; (SS), I.C.S.T.C. on S.C. (TC-S.S.C.P.I.C. (PIC) at I.B.M.R. and H.S.S. Heuristic Recovery of Missing Events in Process Logs. In Proceedings of the IEEE International Conference on Web Services, ICWS 2015; Zhu, H., Miller, J.A., Eds.; Institute of Electrical and Electronics Engineers Inc., 2015; pp. 105–112.
130. Kong, L.; Li, C.; Ge, J.; Li, Z.; Zhang, F.; Luo, B.; Institute for Systems and Technologies of Information, C. and C. (INSTICC) An Efficient Heuristic Method for Repairing Event Logs Independent of Process Models. In Proceedings of the 4th International Conference on Internet of Things, Big Data and Security, IoTBDS 2019; Ramachandran, M., Walters, R., Wills, G., Munoz, V.M., Chang, V., Eds.; SciTePress, 2019; pp. 83–93.
131. Lu, X.; Fahland, D.; van den Biggelaar, F.J.H.M.; van der Aalst, W.M.P. Handling Duplicated Tasks in Process Discovery by Refining Event Labels. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2016**, *9850 LNCS*, 90–107, doi:10.1007/978-3-319-45348-4_6.
132. Dixit, P.M.; Suriadi, S.; Andrews, R.; Wynn, M.T.; ter Hofstede, A.H.M.; Buijs, J.C.A.M.; van der Aalst, W.M.P. Detection and Interactive Repair of Event Ordering Imperfection in Process Logs. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Tallinn, 2018; Vol. 10816, pp. 274–290.
133. Richetti, P.H.P.; Baião, F.A.; Santoro, F.M. Declarative Process Mining: Reducing Discovered Models Complexity by Pre-Processing Event Logs. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2014**, *8659 LNCS*, 400–407, doi:10.1007/978-3-319-10172-9_28.
134. Ekici, B.; Tarhan, A.; Ozsoy, A. Data Cleaning for Process Mining with Smart Contract. In Proceedings of the 4th International Conference on Computer Science and Engineering, UBMK 2019; Institute of Electrical and Electronics Engineers Inc., 2019; pp. 324–329.
135. vanden Broucke, S.K.L.M.; Delvaux, C.; Freitas, J.; Rogova, T.; Vanthienen, J.; Baesens, B. Uncovering the Relationship Between Event Log Characteristics and Process Discovery Techniques. In *Lecture Notes in Business Information Processing*; 2014; Vol. 171, pp. 41–53.
136. Weber, P.; Bordbar, B.; Tino, P. A Framework for the Analysis of Process Mining Algorithms. *IEEE Trans Syst Man Cybern Syst* **2013**, *43*, 303–317, doi:10.1109/TSMCA.2012.2195169.

137. Weber, P.; Bordbar, B.; Tino, P.; Majeed, B. A Framework for Comparing Process Mining Algorithms. In Proceedings of the 2011 IEEE GCC Conference and Exhibition (GCC); IEEE, February 2011; pp. 625–628.
138. Wang, J.; Wong, R.K.; Ding, J.; Guo, Q.; Wen, L. On Recommendation of Process Mining Algorithms. In Proceedings of the 2012 IEEE 19th International Conference on Web Services; IEEE: Honolulu, Hawaii, June 2012; pp. 311–318.
139. Wang, J.; Wong, R.K.; Ding, J.; Guo, Q.; Wen, L. Efficient Selection of Process Mining Algorithms. *IEEE Trans Serv Comput* **2013**, *6*, 484–496, doi:10.1109/TSC.2012.20.
140. Milani, F.; Lashkevich, K.; Maggi, F.M.; Di Francescomarino, C. Process Mining: A Guide for Practitioners. In *Lecture Notes in Business Information Processing*; Guizzardi, R., Ralyté, J., Franch, X., Eds.; Springer, Cham, 2022; Vol. 446, pp. 265–282.
141. Tiwari, A.; Turner, C.J.; Majeed, B. A Review of Business Process Mining: State-of-the-art and Future Trends. *Business Process Management Journal* **2008**, *14*, 5–22, doi:10.1108/14637150810849373.
142. van Dongen, B.F.; Alves de Medeiros, A.K.; Wen, L. Process Mining: Overview and Outlook of Petri Net Discovery Algorithms. In *Lecture Notes in Computer Sciences*; Springer, Berlin, Heidelberg, 2009; Vol. 5460, pp. 225–242.
143. Dianmin Yue; Xiaodan Wu; Haiyan Wang; Junbo Bai A Review of Process Mining Algorithms. In Proceedings of the 2011 International Conference on Business Management and Electronic Information; IEEE: Guangzhou, May 2011; pp. 181–185.
144. van der Aalst, W. Process Mining: Overview and Opportunities. *ACM Trans Manag Inf Syst* **2012**, *3*, 1–17, doi:10.1145/2229156.2229157.
145. Li, H.; Wei, Y.; Liu, L.; Yao, S.W.; Yang, J. Process Mining: Overview and Comparative Analysis of the Mining Algorithms. *Adv Mat Res* **2014**, 989–994, 1924–1929, doi:10.4028/www.scientific.net/AMR.989-994.1924.
146. Martin, N.; Depaire, B.; Caris, A. The Use of Process Mining in a Business Process Simulation Context: Overview and Challenges. In Proceedings of the 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM); IEEE, December 2014; pp. 381–388.
147. Ahmed, R.; Faizan, M.; Burney, A.I. Process Mining in Data Science: A Literature Review. In Proceedings of the 2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS); IEEE, December 2019; pp. 1–9.
148. Erdem, S.; Demirörs, O.; Rabhi, F. Systematic Mapping Study on Process Mining in Agile Software Development. In; 2018; pp. 289–299.
149. Ghazal, M.A.; Ibrahim, O.; Salama, M.A. Educational Process Mining: A Systematic Literature Review. In Proceedings of the 2017 European Conference on Electrical Engineering and Computer Science (EECS); IEEE, November 2017; pp. 198–203.
150. Jokonowo, B.; Claes, J.; Sarno, R.; Rochimah, S. Process Mining in Supply Chains: A Systematic Literature Review. *International Journal of Electrical and Computer Engineering (IJECE)* **2018**, *8*, 4626, doi:10.11591/ijece.v8i6.pp4626-4636.
151. El-Gharib, N.M.; Amyot, D. Process Mining for Cloud-Based Applications: A Systematic Literature Review. In Proceedings of the 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW); IEEE, September 2019; pp. 34–43.
152. Macak, M.; Daubner, L.; Sani, M.F.; Buhnova, B. Cybersecurity Analysis via Process Mining: A Systematic Literature Review. In *Lecture Notes in Computer Science*; Li, B., Ed.; Springer, Cham, 2022; Vol. 13087, pp. 393–407.

153. Maita, A.R.C.; Martins, L.C.; López Paz, C.R.; Rafferty, L.; Hung, P.C.K.; Peres, S.M.; Fantinato, M. A Systematic Mapping Study of Process Mining. *Enterp Inf Syst* **2018**, *12*, 505–549, doi:10.1080/17517575.2017.1402371.
154. Garcia, C. dos S.; Meincheim, A.; Faria Junior, E.R.; Dallagassa, M.R.; Sato, D.M.V.; Carvalho, D.R.; Santos, E.A.P.; Scalabrin, E.E. Process Mining Techniques and Applications – A Systematic Mapping Study. *Expert Syst Appl* **2019**, *133*, 260–295, doi:10.1016/j.eswa.2019.05.003.
155. Yasmin, F.A.; Bukhsh, F.A.; De Alencar Silva, P. Process Enhancement in Process Mining: A Literature Review. In Proceedings of the 8th International Symposium on Data-Driven Process Discovery and Analysis, SIMPDA 2018; Ceur.org: Sevilla, Spain, 2018.
156. Ghasemi, M.; Amyot, D. From Event Logs to Goals: A Systematic Literature Review of Goal-Oriented Process Mining. *Requir Eng* **2020**, *25*, 67–93, doi:10.1007/s00766-018-00308-3.
157. Duan, C.; Wei, Q. Process Mining of Duplicate Tasks: A Systematic Literature Review. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA); IEEE, June 2020; pp. 778–784.
158. Arias, M.; Saavedra, R.; Marques, M.R.; Munoz-Gama, J.; Sepúlveda, M. Human Resource Allocation in Business Process Management and Process Mining. *Management Decision* **2018**, *56*, 376–405, doi:10.1108/MD-05-2017-0476.
159. Ghasemi, M.; Amyot, D. Process Mining in Healthcare: A Systematised Literature Review. *Int J Electron Healthc* **2016**, *9*, 60, doi:10.1504/IJEH.2016.078745.
160. Rojas, E.; Munoz-Gama, J.; Sepúlveda, M.; Capurro, D. Process Mining in Healthcare: A Literature Review. *J Biomed Inform* **2016**, *61*, 224–236, doi:10.1016/j.jbi.2016.04.007.
161. Kurniati, A.P.; Johnson, O.; Hogg, D.; Hall, G. Process Mining in Oncology: A Literature Review. In Proceedings of the 2016 6th International Conference on Information Communication and Management (ICICM); IEEE, October 2016; pp. 291–297.
162. Williams, R.; Rojas, E.; Peek, N.; Johnson, O. Process Mining in Primary Care: A Literature Review. *Stud Health Technol Inform* **2018**, *247*, 376–380.
163. Batista, E.; Solanas, A. Process Mining in Healthcare: A Systematic Review. In Proceedings of the 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA); IEEE, July 2018; pp. 1–6.
164. Farid, N.; De Kamps, M.; Johnson, O. Process Mining in Frail Elderly Care: A Literature Review. In Proceedings of the Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies; SCITEPRESS - Science and Technology Publications, 2019; pp. 332–339.
165. Shanmuga Sundari, M.; Kalyan Nayak, N. Process Mining in Healthcare Systems: A Critical Review and Its Future. *International Journal of Emerging Trends in Engineering Research* **2020**, *8*, 5197–5208, doi:10.30534/ijeter/2020/50892020.
166. Corallo, A.; Lazoi, M.; Striani, F. Process Mining and Industrial Applications: A Systematic Literature Review. *Knowledge and Process Management* **2020**, *27*, 225–233, doi:10.1002/kpm.1630.
167. Chaydy, N.; Madani, A. An Overview of Process Mining and Its Applicability to Complex, Real-Life Scenarios. In Proceedings of the 2019 International Conference on Systems of Collaboration Big Data, Internet of Things & Security (SysCoBioTS); IEEE, December 2019; pp. 1–9.
168. Thiede, M.; Fuerstenau, D.; Bezerra Barquet, A.P. How Is Process Mining Technology Used by Organizations? A Systematic Literature Review of Empirical Studies. *Business Process Management Journal* **2018**, *24*, 900–922, doi:10.1108/BPMJ-06-2017-0148.

169. Dallagassa, M.R.; dos Santos Garcia, C.; Scalabrin, E.E.; Ioshii, S.O.; Carvalho, D.R. Opportunities and Challenges for Applying Process Mining in Healthcare: A Systematic Mapping Study. *J Ambient Intell Humaniz Comput* **2022**, *13*, 165–182, doi:10.1007/s12652-021-02894-7.
170. Ifenthaler, D.; Schumacher, C.; Kuzilek, J. Investigating Students' Use of Self-assessments in Higher Education Using Learning Analytics. *J Comput Assist Learn* **2022**, *39*, 255–268, doi:10.1111/jcal.12744.
171. Werner, M.; Gehrke, N. Identifying the Absence of Effective Internal Controls: An Alternative Approach for Internal Control Audits. *Journal of Information Systems* **2019**, *33*, 205–222, doi:10.2308/isis-52112.
172. Daniel, C.; Gill, T.; Hevner, A.R.; Mullarkey, M. A Deep Neural Network Approach to Tracing Paths in Cybersecurity Investigations.; Fatta, G. Di, Sheng, V., Cuzzocrea, A., Zaniolo, C., Wu, X., Eds.; IEEE Computer Society, 2020; Vol. 2020-November, pp. 472–479.
173. Gibert, K.; Sánchez-Marrè, M.; Codina, V. Choosing the Right Data Mining Technique: Classification of Methods and Intelligent Recommendation. In Proceedings of the 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake; Ottawa, Canada, 2012.
174. Chikohora, T. A Study of the Factors Considered When Choosing an Appropriate Data Mining Algorithm. *International Journal of Soft Computing and Engineering* **2014**, *4*.
175. García, S.; Luengo, J.; Herrera, F. *Data Preprocessing in Data Mining*; Springer International Publishing: Cham, 2015; Vol. 72; ISBN 978-3-319-10246-7.
176. Jiawei, H.; Jian Pei; Micheline Kamber *Data Mining: Concepts and Techniques*; Elsevier, 2012; ISBN 9780123814791.
177. Shiferaw, R.; Bogale, A.; Debela, K. Implementing Research Methods with Confidence: A Review of Research Methodology: A Step-by-Step Guide for Beginners. *The Qualitative Report* **2022**, *27*, doi:10.46743/2160-3715/2022.6024.
178. Tierney, A.J. Research Design in Social Research. *Int J Nurs Stud* **2002**, *39*, 669–670, doi:10.1016/S0020-7489(01)00040-2.
179. Michell, J. Measurement Scales and Statistics: A Clash of Paradigms. *Psychol Bull* **1986**, *100*, 398–407, doi:10.1037/0033-2909.100.3.398.
180. Wackerly Dennis; Mendenhall III William; Scheaffer Richard *Mathematical Statistics with Applications*; 2008;
181. Kitchenham, B. *Procedures for Undertaking Systematic Reviews*; 2004;
182. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, doi:10.1136/bmj.n71.
183. van Eck, N.J.; Waltman, L. Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping. *Scientometrics* **2010**, *84*, 523–538, doi:10.1007/s11192-009-0146-3.
184. Hoffman, J.I.E. Hypergeometric Distribution. In *Biostatistics for Medical and Biomedical Practitioners*; Elsevier, 2015; pp. 179–182.
185. Groves Floyd J Fowler, R.M.; Mick Couper James Lepkowski Eleanor Singer Roger Tourangeau, J.M.; Walter Shewhart, by A.; Wilks, S.S. *Survey Methodology*; 2011;
186. Jebb, A.T.; Ng, V.; Tay, L. A Review of Key Likert Scale Development Advances: 1995–2019. *Front Psychol* **2021**, *12*, doi:10.3389/fpsyg.2021.637547.
187. Etikan, I. Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics* **2016**, *5*, 1, doi:10.11648/j.ajtas.20160501.11.

188. Campbell, S.; Greenwood, M.; Prior, S.; Shearer, T.; Walkem, K.; Young, S.; Bywaters, D.; Walker, K. Purposive Sampling: Complex or Simple? Research Case Examples. *Journal of Research in Nursing* **2020**, *25*, 652–661, doi:10.1177/1744987120927206.
189. Palinkas, L.A.; Horwitz, S.M.; Green, C.A.; Wisdom, J.P.; Duan, N.; Hoagwood, K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research* **2015**, *42*, 533–544, doi:10.1007/s10488-013-0528-y.
190. Elhadjamor, E.A.; Ghannouchi, S.A. Analyze in Depth Health Care Business Process and Key Performance Indicators Using Process Mining.; Cruz-Cunha, M.M., Varajao, J.E., Martinho, R., Rijo, R., Peres, E., Domingos, D., Eds.; Elsevier B.V., 2019; Vol. 164, pp. 610–617.
191. Benevento, E.; Dixit, P.M.; Sani, M.F.; Aloini, D.; van der Aalst, W.M.P. Evaluating the Effectiveness of Interactive Process Discovery in Healthcare: A Case Study 2019, *362 LNBIP*, 508–519.
192. Aguirre, J.A.; Torres, A.C.; Pescoran, M.E.; Mayorga, S.A. Evaluation of Operational Process Variables in Healthcare Using Process Mining and Data Visualization Techniques.; Latin American and Caribbean Consortium of Engineering Institutions, 2019; Vol. 2019-July.
193. Araghi, S.N.; Fontanili, F.; Lamine, E.; Salatge, N.; Lesbegueries, J.; Pouyade, S.R.; Benaben, F. Evaluating the Process Capability Ratio of Patients' Pathways by the Application of Process Mining, SPC and RTLS.; Moucek, R., Fred, A., Gamboa, H., Eds.; SciTePress, 2019; pp. 302–309.
194. Xu, H.; Pang, J.; Yang, X.; Yu, J.; Li, X.; Zhao, D. Modeling Clinical Activities Based on Multi-Perspective Declarative Process Mining with OpenEHR's Characteristic. *BMC Med Inform Decis Mak* **2020**, *20*, 303, doi:10.1186/s12911-020-01323-7.
195. Villamor, D.A.R.; Pulmano, C.E.; Estuar, M.R.J.E. Understanding Adoption of Electronic Medical Records: Application of Process Mining for Health Worker Behavior Analysis.; Association for Computing Machinery, 2020; pp. 98–104.
196. Agostinelli, S.; Covino, F.; D'Agnesse, G.; Crea, C. De; Leotta, F.; Marrella, A. Supporting Governance in Healthcare through Process Mining: A Case Study. *IEEE Access* **2020**, *8*, 186012–186025, doi:10.1109/ACCESS.2020.3030318.
197. Stefanini, A.; Aloini, D.; Benevento, E.; Dulmin, R.; Mininno, V. A Process Mining Methodology for Modeling Unstructured Processes. *Knowledge and Process Management* **2020**, *27*, 294–310, doi:10.1002/kpm.1649.
198. Pang, J.; Xu, H.; Ren, J.; Yang, J.; Li, M.; Lu, D.; Zhao, D. Process Mining Framework with Time Perspective for Understanding Acute Care: A Case Study of AIS in Hospitals. *BMC Med Inform Decis Mak* **2021**, *21*, 354, doi:10.1186/s12911-021-01725-1.
199. Bano, D.; Michael, J.; Rumpe, B.; Varga, S.; Weske, M. Process-Aware Digital Twin Cockpit Synthesis from Event Logs. *J Comput Lang* **2022**, *70*, doi:10.1016/j.cola.2022.101121.
200. Araghi, S.N.; Fontanili, F.; Lamine, E.; Okongwu, U.; Benaben, F. Stable Heuristic Miner: Applying Statistical Stability to Discover the Common Patient Pathways from Location Event Logs. *Intelligent Systems with Applications* **2022**, *14*, doi:10.1016/j.iswa.2022.200071.
201. Fontenla-Seco, Y.; Lama, M.; González-Salvado, V.; Peña-Gil, C.; Bugarín-Diz, A. A Framework for the Automatic Description of Healthcare Processes in Natural Language: Application in an Aortic Stenosis Integrated Care Process. *J Biomed Inform* **2022**, *128*, doi:10.1016/j.jbi.2022.104033.
202. Erdogan, T.G.; Tarhan, A.K. Multi-Perspective Process Mining for Emergency Process. *Health Informatics J* **2022**, *28*, doi:10.1177/14604582221077195.

203. Park, G.; Adams, J.N.; van der Aalst, W.M.P. OPerA: Object-Centric Performance Analysis. In: 2022; pp. 281–292.
204. Pegoraro, M.; Narayana, M.B.S.; Benevento, E.; van der Aalst, W.M.P.; Martin, L.; Marx, G. Analyzing Medical Data with Process Mining: A COVID-19 Case Study 2022, *444 LNBIP*, 39–44.
205. Rashed, A.-H.M.; El-Attar, N.E.; Abdelminaam, D.S.; Abdelfatah, M. Analysis the Patients' Careflows Using Process Mining. *PLoS One* **2023**, *18*, e0281836, doi:10.1371/journal.pone.0281836.
206. Rismanchian, F.; Kassani, S.H.; Shavarani, S.M.; Lee, Y.H. A Data-Driven Approach to Support the Understanding and Improvement of Patients' Journeys: A Case Study Using Electronic Health Records of an Emergency Department. *Value in Health* **2023**, *26*, 18–27, doi:10.1016/j.jval.2022.04.002.
207. Marazza, F.; Bukhsh, F.A.; Vijlbrief, O.; Geerdink, J.; Pathak, S.; van Keulen, M.; Seifert, C. Comparing Process Models for Patient Populations: Application in Breast Cancer Care 2019, *362 LNBIP*, 496–507.
208. Canjels, K.F.; Imkamp, M.S. V; Boymans, T.A.E.J.; Vanwersch, R.J.B. Unraveling and Improving the Interorganizational Arthrosis Care Process at Maastricht UMC+: An Illustration of an Innovative, Combined Application of Data and Process Mining.; vom Brocke, J., Mendling, J., Rosemann, M., Eds.; CEUR-WS, 2019; Vol. 2428, pp. 178–189.
209. Marazza, F.; Bukhsh, F.A.; Geerdink, J.; Vijlbrief, O.; Pathak, S.; van Keulen, M.; Seifert, C. Automatic Process Comparison for Subpopulations: Application in Cancer Care. *Int J Environ Res Public Health* **2020**, *17*, 1–23, doi:10.3390/ijerph17165707.
210. Dallagassa, M.R.; Iachecen, F.; Furlan, L.H.P.; Ioshii, S.O.; Carvalho, D.R. Applying Process Mining in Health Technology Assessment. *Health Technol (Berl)* **2022**, *12*, 931–941, doi:10.1007/s12553-022-00692-5.
211. Cho, M.; Song, M.; Yoo, S.; Reijers, H.A. An Evidence-Based Decision Support Framework for Clinician Medical Scheduling. *IEEE Access* **2019**, *7*, 15239–15249, doi:10.1109/ACCESS.2019.2894116.
212. Dogan, O. Process Mining Based on Patient Waiting Time: An Application in Health Processes. *International Journal of Web Information Systems* **2022**, *18*, 240–254, doi:10.1108/IJWIS-02-2022-0027.
213. Francescomarino, C. Di; Dumas, M.; Maggi, F.M.; Teinemaa, I. Clustering-Based Predictive Process Monitoring. *IEEE Trans Serv Comput* **2019**, *12*, 896–909, doi:10.1109/TSC.2016.2645153.
214. Amantea, I.A.; Sulis, E.; Boella, G.; Marinello, R.; Bianca, D.; Brunetti, E.; Bo, M.; Fernandez-Llatas, C. A Process Mining Application for the Analysis of Hospital-at-Home Admissions 2020, *270*, 522–526.
215. Pijnenborg, P.; Verhoeven, R.; Firat, M.; Laarhoven, H. V; Genga, L. Towards Evidence-Based Analysis of Palliative Treatments for Stomach and Esophageal Cancer Patients: A Process Mining Approach.; Ciccio, C. Di, Francescomarino, C. Di, Soffer, P., Eds.; Institute of Electrical and Electronics Engineers Inc., 2021; pp. 136–143.
216. Aringhieri, R.; Boella, G.; Brunetti, E.; Caro, L. Di; Francescomarino, C. Di; Dragoni, M.; Ferrod, R.; Ghidini, C.; Marinello, R.; Ronzani, M.; et al. Towards the Application of Process Mining for Supporting the Home Hospitalization Service.; Marrella, A., Dupre, D.T., Eds.; CEUR-WS, 2021; Vol. 2952, pp. 33–38.

217. Kurniati, A.P.; McInerney, C.; Zucker, K.; Hall, G.; Hogg, D.; Johnson, O. Using a Multi-Level Process Comparison for Process Change Analysis in Cancer Pathways. *Int J Environ Res Public Health* **2020**, *17*, 1–16, doi:10.3390/ijerph17197210.
218. Stefanini, A.; Aloini, D.; Benevento, E.; Dulmin, R.; Mininno, V. A Data-Driven Methodology for Supporting Resource Planning of Health Services. *Socioecon Plann Sci* **2020**, *70*, doi:10.1016/j.seps.2019.100744.
219. Li, H.; Liu, C.; Zeng, Q.; He, H.; Ren, C.; Wang, L.; Cheng, F. Mining Emergency Event Logs to Support Resource Allocation. *IEICE Trans Inf Syst* **2021**, *E104D*, 1651–1660, doi:10.1587/transinf.2021EDP7029.
220. Tavazzi, E.; Gerard, C.L.; Michielin, O.; Wicky, A.; Gatta, R.; Cuendet, M.A. A Process Mining Approach to Statistical Analysis: Application to a Real-World Advanced Melanoma Dataset 2021, *406 LNBIP*, 291–304.
221. Matthanawongsakorn, C.; Saguansakdiyotin, N.; Porouhan, P.; Arpasat, P.; Premochaiswadi, W. Applying Process Mining to Investigate the Relation between Food Purchase Behavior and Children's Weight Based on the Food Digital Cards.; IEEE Computer Society, 2019; Vol. 2019-November.
222. Rojas, E.; Capurro, D. Characterization of Drug Use Patterns Using Process Mining and Temporal Abstraction Digital Phenotyping 2019, *342*, 187–198.
223. Beerepoot, I.; Lu, X.; van de Weerd, I.; Reijers, H.A. Seeing the Signs of Workarounds: A Mixed-Methods Approach to the Detection of Nurses' Process Deviations.; Bui, T.X., Ed.; IEEE Computer Society, 2021; Vol. 2020-January, pp. 3763–3772.
224. Rojas, E.; Cifuentes, A.; Burattin, A.; Munoz-Gama, J.; Sepúlveda, M.; Capurro, D. Analysis of Emergency Room Episodes Duration Through Process Mining. In; 2019; pp. 251–263.
225. Arias, M.; Rojas, E.; Aguirre, S.; Cornejo, F.; Munoz-Gama, J.; Sepúlveda, M.; Capurro, D. Mapping the Patient's Journey in Healthcare through Process Mining. *Int J Environ Res Public Health* **2020**, *17*, 1–16, doi:10.3390/ijerph17186586.
226. Cho, M.; Song, M.; Park, J.; Yeom, S.R.; Wang, I.J.; Choi, B.K. Process Mining-Supported Emergency Room Process Performance Indicators. *Int J Environ Res Public Health* **2020**, *17*, 1–20, doi:10.3390/ijerph17176290.
227. Duma, D.; Aringhieri, R. An Ad Hoc Process Mining Approach to Discover Patient Paths of an Emergency Department. *Flex Serv Manuf J* **2018**, *32*, 6–34, doi:10.1007/s10696-018-9330-1.
228. Esiefarienrhe, B.M.; Omolewa, I.D. Application of Process Mining to Medical Billing Using L* Life Cycle Model.; Institute of Electrical and Electronics Engineers Inc., 2021.
229. Kusuma, G.; Kurniati, A.; McInerney, C.D.; Hall, M.; Gale, C.P.; Johnson, O. Process Mining of Disease Trajectories in MIMIC-III: A Case Study 2021, *406 LNBIP*, 305–316.
230. Singh, S.; Verma, R.; Koul, S. A Collaborative Method for Simultaneous Operations: Case of an Eye Clinic. *OPSEARCH* **2022**, *59*, 711–731, doi:10.1007/s12597-021-00513-9.
231. Kropp, T.; Faeghi, S.; Lennerts, K. Evaluation of Patient Transport Service in Hospitals Using Process Mining Methods: Patients' Perspective. *Int J Health Plann Manage* **2023**, *38*, 430–456, doi:10.1002/hpm.3593.
232. Tsai, E.R.; Tintu, A.N.; Boucherie, R.J.; de Rijke, Y.B.; Schotman, H.H.M.; Demirtas, D. Characterization of Laboratory Flow and Performance for Process Improvements via Application of Process Mining. *Appl Clin Inform* **2023**, *14*, 144–152, doi:10.1055/a-1996-8479.
233. Mohammadi, F.; Kazempourian, S.; Vanani, I.R. Process Mining Approach to Performance Analysis and Bottleneck Finding in Electronic Processes (Case Study: The Billing Process of

- Hospital Services). *International Journal of Process Management and Benchmarking* **2023**, *13*, 212–232, doi:10.1504/IJPMB.2021.10040004.
234. Elkhovskaya, L.O.; Kshenin, A.D.; Balakhontceva, M.A.; Ionov, M. V; Kovalchuk, S. V Extending Process Discovery with Model Complexity Optimization and Cyclic States Identification: Application to Healthcare Processes. *Algorithms* **2023**, *16*, doi:10.3390/a16010057.
 235. Shin, S.C.; Kim, S.Y.; Noh, C.M.; Lee, S.S.; Lee, J.C. Manufacturing Process Improvement of Offshore Plant: Process Mining Technique and Case Study. *Ocean Systems Engineering* **2019**, *9*, 329–347, doi:10.12989/ose.2019.9.3.329.
 236. Altan, Z.; Birgün, S. Using Process Mining Approach for Machining Operations 2020, 452–464.
 237. Birk, A.; Wilhelm, Y.; Dreher, S.; Flack, C.; Reimann, P.; Gröger, C. A Real-World Application of Process Mining for Data-Driven Analysis of Multi-Level Interlinked Manufacturing Processes.; Mourtzis, D., Ed.; Elsevier B.V., 2021; Vol. 104, pp. 417–422.
 238. Ruschel, E.; Loures, E.D.F.R.; Santos, E.A.P. Performance Analysis and Time Prediction in Manufacturing Systems. *Comput Ind Eng* **2021**, *151*, doi:10.1016/j.cie.2020.106972.
 239. Lashkevich, K.; Milani, F.; Chapela-Campa, D.; Dumas, M. Data-Driven Analysis of Batch Processing Inefficiencies in Business Processes. In Proceedings of the Lecture Notes in Business Information Processing; Springer Science and Business Media Deutschland GmbH, 2022; Vol. 446 LNBIP, pp. 231–247.
 240. Kumbhar, M.; Ng, A.H.C.; Bandaru, S. Bottleneck Detection Through Data Integration, Process Mining and Factory Physics-Based Analytics.; Ng, A.H.C., Syberfeldt, A., Hogberg, D., Holm, M., Eds.; IOS Press BV, 2022; Vol. 21, pp. 737–748.
 241. Scheibel, B.; Rinderle-Ma, S. Decision Mining with Time Series Data Based on Automatic Feature Generation 2022, *13295 LNCS*, 3–18.
 242. Galanti, R.; de Leoni, M.; Navarin, N.; Marazzi, A. Object-Centric Process Predictive Analytics. *Expert Syst Appl* **2023**, *213*, doi:10.1016/j.eswa.2022.119173.
 243. Lugaresi, G.; Ciappina, A.D.; Rossi, M.; Matta, A. Exploiting a Combined Process Mining Approach to Enhance the Discovery and Analysis of Support Processes in Manufacturing. *Int J Comput Integr Manuf* **2023**, *36*, 169–189, doi:10.1080/0951192X.2022.2090024.
 244. Calderón-Ruiz, G.; Luque, D.F. Process Mining: The First Successful Peruvian Case.; Petrie, M.M.L., Texier, J., Pena, A., Vilorio, J.A.S., Eds.; Latin American and Caribbean Consortium of Engineering Institutions, 2022; Vol. 2022-July.
 245. Wu, Q.; He, Z.; Wang, H.; Wen, L.; Yu, T. A Business Process Analysis Methodology Based on Process Mining for Complaint Handling Service Processes. *Applied Sciences (Switzerland)* **2019**, *9*, doi:10.3390/app9163313.
 246. Abonyi, J.; Dorgo, G. Process Mining in Production Systems.; Institute of Electrical and Electronics Engineers Inc., 2019; pp. 267–270.
 247. Bántay, L.; Dörgö, G.; Tandari, F.; Abonyi, J. Simultaneous Process Mining of Process Events and Operator Actions for Alarm Management. *Complexity* **2022**, *2022*, doi:10.1155/2022/8670154.
 248. Halawa, F.; Madathil, S.C.; Khasawneh, M.T. Integrated Framework of Process Mining and Simulation–Optimization for Pod Structured Clinical Layout Design. *Expert Syst Appl* **2021**, *185*, doi:10.1016/j.eswa.2021.115696.
 249. Rudnitskaia, J.; Venkatachalam, H.S.; Essmann, R.; Hruska, T.; Colombo, A.W. Screening Process Mining and Value Stream Techniques on Industrial Manufacturing Processes: Process

- Modelling and Bottleneck Analysis. *IEEE Access* **2022**, *10*, 24203–24214, doi:10.1109/ACCESS.2022.3152211.
250. Bahaweres, R.B.; Amna, H.; Nurnaningsih, D. Improving Purchase to Pay Process Efficiency with RPA Using Fuzzy Miner Algorithm in Process Mining.; Institute of Electrical and Electronics Engineers Inc., 2022; pp. 1483–1488.
 251. Rigger, E.; Vosgien, T.; Bitrus, S.; Szabo, P.; Eynard, B. Enterprise Architecture Method for Continuous Improvement of PLM Based on Process Mining 2020, *594*, 563–575.
 252. Nafasa, P.; Waspada, I.; Bahtiar, N.; Wibowo, A. Implementation of Alpha Miner Algorithm in Process Mining Application Development for Online Learning Activities Based on MOODLE Event Log Data.; Institute of Electrical and Electronics Engineers Inc., 2019.
 253. Dolak, R. Using Process Mining Techniques to Discover Student’s Activities, Navigation Paths, and Behavior in LMS Moodle 2019, *11937 LNCS*, 129–138.
 254. Sonnenberg, C.; Bannert, M. Using Process Mining to Examine the Sustainability of Instructional Support: How Stable Are the Effects of Metacognitive Prompting on Self-Regulatory Behavior? *Comput Human Behav* **2019**, *96*, 259–272, doi:10.1016/j.chb.2018.06.003.
 255. Real, E.M.; Pimentel, E.P.; Oliveira, L.V. De; Braga, J.C.; Stiubiener, I. Educational Process Mining for Verifying Student Learning Paths in an Introductory Programming Course.; Institute of Electrical and Electronics Engineers Inc., 2020; Vol. 2020-October.
 256. Piotrkowicz, A.; Wang, K.; Hallam, J.; Dimitrova, V. Data-Driven Exploration of Engagement with Workplace-Based Assessment in the Clinical Skills Domain. *Int J Artif Intell Educ* **2021**, *31*, 1022–1052, doi:10.1007/s40593-021-00264-0.
 257. Hachicha, W.; Ghorbel, L.; Champagnat, R.; Zayani, C.A.; Amous, I. Using Process Mining for Learning Resource Recommendation: A Moodle Case Study.; Watrobski, J., Salabun, W., Toro, C., Zanni-Merk, C., Howlett, R.J., Jain, L.C., Jain, L.C., Eds.; Elsevier B.V., 2021; Vol. 192, pp. 853–862.
 258. Okoye, K.; Hosseini, S. Educational Process Intelligence: A Process Mining Approach and Model Analysis 2021, *1180 AISC*, 201–212.
 259. AlQaheri, H.; Panda, M. An Education Process Mining Framework: Unveiling Meaningful Information for Understanding Students’ Learning Behavior and Improving Teaching Quality. *Information* **2022**, *13*, doi:10.3390/info13010029.
 260. Dorrer, M.; Dorrer, A. Generation of Agent Simulation Models by Using Process Mining Methods on the Example of E-Learning Process.; Kovalev, I. V, Voroshilova, A.A., Borisova, E.A., Eds.; Institute of Physics Publishing, 2019; Vol. 1399.
 261. Salazar-Fernandez, J.P.; Sepúlveda, M.; Munoz-Gama, J. Influence of Student Diversity on Educational Trajectories in Engineering High-Failure Rate Courses That Lead to Late Dropout.; Ashmawy, A.K., Schreiter, S., Eds.; IEEE Computer Society, 2019; Vol. April-2019, pp. 607–616.
 262. Leno, V.; Augusto, A.; Dumas, M.; Rosa, M. La; Maggi, F.M.; Polyvyanyy, A. Discovering Data Transfer Routines from User Interaction Logs. *Inf Syst* **2022**, *107*, doi:10.1016/j.is.2021.101916.
 263. Dorrer, M.G.; Popov, A.A. Application of Process Mining Technology to Analyze the Correspondence of Life Cycles of Information Resources to a Generalized Model.; Kovalev, I. V, Voroshilova, A.A., Testoyedov, N.A., Eds.; IOP Publishing Ltd, 2020; Vol. 1679.
 264. Majumdar, R.; Bakilapadavu, G.; Li, J.; Chen, M.R.A.; Flanagan, B.; Ogata, H. Analytics of Open-Book Exams with Interaction Traces in a Humanities Course.; Rodrigo, M.M.T., Iyer, S., Mitrovic, A., Cheng, H.N.H., Kohen-Vacs, D., Matuk, C., Palalas, A., Rajenran, R., Seta, K., Wang, J., Eds.; Asia-Pacific Society for Computers in Education, 2021; Vol. 1, pp. 352–361.

265. Er, E.; Villa-Torrano, C.; Dimitriadis, Y.; Gasevic, D.; Bote-Lorenzo, M.L.; Asensio-Pérez, J.I.; Gómez-Sánchez, E.; Monés, A.M. Theory-Based Learning Analytics to Explore Student Engagement Patterns in a Peer Review Activity.; Association for Computing Machinery, 2021; pp. 196–206.
266. Doğan, O.; Tirpan, E.C. Process Mining Methodology for Digital Processes under Smart Campus Concept. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi* **2022**, doi:10.35193/bseufbd.1162284.
267. Rahmawati, R.; Andreswari, R.; Fauzi, R. Analysis and Exploratory of Lecture Preparation Process to Improve the Conformance Using Process Mining.; Paul, R., Ed.; Institute of Electrical and Electronics Engineers Inc., 2022; pp. 461–466.
268. Emara, M.; Hutchins, N.M.; Grover, S.; Snyder, C.; Biswas, G. Examining Student Regulation of Collaborative, Computational, Problem-Solving Processes in Openended Learning Environments. *Journal of Learning Analytics* **2021**, 8, 49–74, doi:10.18608/JLA.2021.7230.
269. Shynkarenko, V.; Zhevaho, O. Application of Constructive Modeling and Process Mining Approaches to the Study of Source Code Development in Software Engineering Courses. *Journal of Communications Software and Systems* **2021**, 17, 342–349, doi:10.24138/JCOMSS-2021-0046.
270. Meengoen, P.; Porouhan, P.; Kungcharoen, K.; Palangsantikul, P.; Arpasat, P.; Premchaiswadi, W. Analysis of Materials and Supplies Withdrawal Process in Educational Institutions with Fuzzy Miner Technique.; IEEE Computer Society, 2020; Vol. 2020-November.
271. Salazar-Fernandez, J.P.; Sepúlveda, M.; Munoz-Gama, J. Describing Educational Trajectories of Engineering Students in Individual High-Failure Rate Courses That Lead to Late Dropout.; Scheihing, E., Guerra, J., Henriquez, V., Olivares, C., Munoz-Merino, P.J., Eds.; CEUR-WS, 2019; Vol. 2425, pp. 39–48.
272. Fernández-Cerero, D.; Varela-Vaca, Á.J.; Fernández-Montes, A.; Gómez-López, M.T.; Álvarez-Bermejo, J.A. Measuring Data-Centre Workflows Complexity through Process Mining: The Google Cluster Case. *J Supercomput* **2019**, 76, 2449–2478, doi:10.1007/s11227-019-02996-2.
273. Taskesenlioglu, S.; Ozkan, N.; Erdogan, T.G. Identifying Possible Improvements of Software Development Life Cycle (SDLC) Process of a Bank by Using Process Mining. *International Journal of Software Engineering and Knowledge Engineering* **2022**, 32, 525–552, doi:10.1142/S0218194022400010.
274. Nogueira, A.F.; Zenha-Rela, M. Monitoring a CI/CD Workflow Using Process Mining. *SN Comput Sci* **2021**, 2, doi:10.1007/s42979-021-00830-2.
275. Azumah, K.K.; Maciel, P.R.M.; Sorensen, L.T.; Kosta, S. Modeling and Simulating a Process Mining-Influenced Load-Balancer for the Hybrid Cloud. *IEEE Transactions on Cloud Computing* **2022**, 1, doi:10.1109/TCC.2022.3177668.
276. Raj, V.; Chander, G.P. Monitoring of Microservices Architecture Based Applications Using Process Mining.; Institute of Electrical and Electronics Engineers Inc., 2022; pp. 486–494.
277. Ramadan, S.; Baqapuri, H.I.; Roecher, E.; Mathiak, K. Process Mining of Logged Gaming Behavior.; Institute of Electrical and Electronics Engineers Inc., 2019; pp. 57–64.
278. Trabelsi, M.; Suire, C.; Morcos, J.; Champagnat, R. User-Centred Application for Modeling Journeys in Digital Libraries.; Downie, J.S., McKay, D., Suleman, H., Nichols, D.M., Poursardar, F., Eds.; Institute of Electrical and Electronics Engineers Inc., 2021; Vol. 2021-September, pp. 332–333.
279. Hobeck, R.; Klinkmüller, C.; Bandara, H.M.N.D.; Weber, I.; van der Aalst, W.M.P. Process Mining on Blockchain Data: A Case Study of Augur 2021, *12875 LNCS*, 306–323.

280. Rojo, J.; Garcia-Alonso, J.; Berrocal, J.; Hernández, J.; Murillo, J.M.; Canal, C. SOWCompact: A Federated Process Mining Method for Social Workflows. *Inf Sci (N Y)* **2022**, *595*, 18–37, doi:10.1016/j.ins.2022.02.035.
281. Hemmer, A.; Abderrahim, M.; Badonnel, R.; Francois, J.; Chrisment, I. Comparative Assessment of Process Mining for Supporting IoT Predictive Security. *IEEE Transactions on Network and Service Management* **2021**, *18*, 1092–1103, doi:10.1109/TNSM.2020.3038172.
282. Macak, M.; Vaclavek, R.; Kusnirakova, D.; Matulevicius, R.; Buhnova, B. Scenarios for Process-Aware Insider Attack Detection in Manufacturing.; Association for Computing Machinery, 2022.
283. Rodriguez, M.; Betarte, G.; Calejari, D. A Process Mining-Based Approach for Attacker Profiling.; Institute of Electrical and Electronics Engineers Inc., 2021; pp. 425–429.
284. Bruno, M.; Ibañez, P.; Techera, T.; Calejari, D.; Betarte, G. Exploring the Application of Process Mining Techniques to Improve Web Application Security.; Institute of Electrical and Electronics Engineers Inc., 2021.
285. Özdağoglu, G.; Kavuncubaşı, E. Monitoring the Software Bug-Fixing Process through the Process Mining Approach. *Journal of Software: Evolution and Process* **2019**, *31*, doi:10.1002/smr.2162.
286. Ardimento, P.; Bernardi, M.L.; Cimitile, M. Malware Phylogeny Analysis Using Data-Aware Declarative Process Mining.; Castellano, G., Castiello, C., Mencar, C., Eds.; Institute of Electrical and Electronics Engineers Inc., 2020; Vol. 2020-May.
287. Gupta, M.; Agarwal, P.; Tater, T.; Dechu, S.; Serebrenik, A. Analyzing Comments in Ticket Resolution to Capture Underlying Process Interactions. In *Business Process Management Workshops*; 2020; pp. 219–231.
288. Ramos-Gutiérrez, B.; Varela-Vaca, Á.J.; Galindo, J.A.; Gómez-López, M.T.; Benavides, D. Discovering Configuration Workflows from Existing Logs Using Process Mining. *Empir Softw Eng* **2021**, *26*, doi:10.1007/s10664-020-09911-x.
289. González, L.; Delgado, A.; Canaparo, J.; Gambetta, F. Evaluation of Compliance Requirements for Collaborative Business Process with Process Mining and a Model of Generic Compliance Controls. *CLEI Electronic Journal (CLEIej)* **2022**, *25*, 71–722, doi:10.19153/CLEIEJ.25.2.7.
290. Houdt, G. Van; Depaire, B.; Martin, N. Root Cause Analysis in Process Mining with Probabilistic Temporal Logic 2022, *433 LNBIP*, 73–84.
291. Terragni, A.; Hassani, M. Optimizing Customer Journey Using Process Mining and Sequence-Aware Recommendation.; Association for Computing Machinery, 2019; Vol. Part F147772, pp. 57–65.
292. Dogan, O.; Fernandez-Llatas, C.; Oztaysi, B. Process Mining Application for Analysis of Customer's Different Visits in a Shopping Mall 2020, *1029*, 151–159.
293. Ocaña, M.; Llamazares, Á.; Revenga, P.A.; García-Garrido, M.A.; Hernández, N.; Álvarez, P.; Fabra, J.; Chapela-Campa, D.; Mucientes, M.; Lama, M.; et al. Estimation of Customer Activity Patterns in Open Malls by Means of Combining Localization and Process Mining Techniques 2021, *1285*, 30–43.
294. Bitkowska, A.; Sliż, P.; Tenbrink, C.; Piasecka, A. Application of Process Mining on the Example of an Authorized Passenger Car Service Station in Poland. *Foundations of Management* **2020**, *12*, 125–136, doi:10.2478/fman-2020-0010.
295. Unger, A.J.; Neto, J.F. dos S.; Fantinato, M.; Peres, S.M.; Trecenti, J.; Hirota, R. Process Mining-Enabled Jurimetrics. In *Proceedings of the Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*; ACM: New York, NY, USA, June 21 2021; pp. 240–244.

296. Halaška, M.; Šperka, R. TDABC and Estimation of Time Drivers Using Process Mining 2021, *241*, 489–499.
297. Saeedi, E.; Safi Esfahani, F. Time and Structural Anomalies Detection in Business Processes Using Process Mining. *International Journal of Business Intelligence and Data Mining* **2019**, *1*, 1, doi:10.1504/IJBIDM.2019.10021265.
298. Šperka, R.; Halaška, M. The Performance Assessment Framework (PPAFR) for RPA Implementation in a Loan Application Process Using Process Mining. *Information Systems and e-Business Management* **2022**, doi:10.1007/s10257-022-00602-2.
299. Poppe, E.; Pika, A.; Wynn, M.T.; Eden, R.; Andrews, R.; ter Hofstede, A.H.M. Extracting Best-Practice Using Mixed-Methods: Insights and Recommendations from a Case Study in Insurance Claims Processing. *Business and Information Systems Engineering* **2021**, *63*, 637–651, doi:10.1007/s12599-021-00698-9.
300. Becker, M.; Buchkremer, R. A Practical Process Mining Approach for Compliance Management. *Journal of Financial Regulation and Compliance* **2019**, *27*, 464–478, doi:10.1108/JFRC-12-2018-0163.
301. Stephan, S.; Lahann, J.; Fettke, P. A Case Study on the Application of Process Mining in Combination with Journal Entry Tests for Financial Auditing.; Bui, T.X., Ed.; IEEE Computer Society, 2021; Vol. 2020-January, pp. 5718–5727.
302. Tridalestari, F.A.; Mustafid, M.; Warsito, B.; Wibowo, A.; Prasetyo, H.N. Analysis of E-Commerce Process in the Downstream Section of Supply Chain Management Based on Process and Data Mining. *Ingenierie des Systemes d'Information* **2022**, *27*, 81–91, doi:10.18280/isi.270110.
303. Maddah, N.; Roghanian, E. Data-Driven Performance Management of Business Units Using Process Mining and DEA: Case Study of an Iranian Chain Store. *International Journal of Productivity and Performance Management* **2023**, *72*, 550–575, doi:10.1108/IJPPM-10-2020-0562.
304. Montianrat, S.; Arpasat, P.; Kungcharoen, K.; Porouhan, P.; Palangsantikul, P.; Premchaiswadi, W. Analysis of Warehouse Operations Using Process Mining Techniques: A Case Study of Scientific Laboratory Equipment and Facilities.; IEEE Computer Society, 2020; Vol. 2020-November.
305. Gunnarsson, B.R.; vanden Broucke, S.K.L.M.; De Weerd, J. Predictive Process Monitoring in Operational Logistics: A Case Study in Aviation. In; 2019; pp. 250–262.
306. Prathama, F.; Yahya, B.N.; Harjono, D.D.; Mahendrawathi, E.R. Trace Clustering Exploration for Detecting Sudden Drift: A Case Study in Logistic Process.; Younus, A., Ed.; Elsevier B.V., 2019; Vol. 161, pp. 1122–1130.
307. Brzychczy, E.; Gackowiec, P.; Liebetrau, M. Data Analytic Approaches for Mining Process Improvement—Machinery Utilization Use Case. *Resources* **2020**, *9*, 17, doi:10.3390/resources9020017.
308. Rashid, K.M.; Louis, J. Process Discovery and Conformance Checking in Modular Construction Using RFID and Process Mining. In Proceedings of the Construction Research Congress 2020; American Society of Civil Engineers: Reston, VA, November 9 2020; pp. 640–648.
309. Pan, Y.; Zhang, L. Automated Process Discovery from Event Logs in BIM Construction Projects. *Autom Constr* **2021**, *127*, doi:10.1016/j.autcon.2021.103713.
310. Rashid, K.M.; Louis, J. Integrating Process Mining with Discrete-Event Simulation for Dynamic Productivity Estimation in Heavy Civil Construction Operations. *Algorithms* **2022**, *15*, doi:10.3390/a15050173.

311. Kropp, T.; Bombeck, A.; Lennerts, K. An Approach to Data Driven Process Discovery in the Cost Estimation Process of a Construction Company.; Feng, C., Linner, T., Brilakis, I., Eds.; International Association for Automation and Robotics in Construction (IAARC), 2021; Vol. 2021-November, pp. 893–900.
312. Pika, A.; ter Hofstede, A.H.M.; Perrons, R.K.; Grossmann, G.; Stumptner, M.; Cooley, J. Using Big Data to Improve Safety Performance: An Application of Process Mining to Enhance Data Visualisation. *Big Data Research* **2021**, *25*, doi:10.1016/j.bdr.2021.100210.
313. Andrews, R.; Wynn, M.T.; Vallmuur, K.; Hofstede, A.H.M. Ter; Bosley, E. A Comparative Process Mining Analysis of Road Trauma Patient Pathways. *Int J Environ Res Public Health* **2020**, *17*, doi:10.3390/ijerph17103426.
314. Chanifah, S.; Andreswari, R.; Fauzi, R. Analysis of Student Learning Pattern in Learning Management System (LMS) Using Heuristic Mining a Process Mining Approach.; Institute of Electrical and Electronics Engineers Inc., 2021; pp. 121–125.
315. Goel, K.; Leemans, S.J.J.; Wynn, M.T.; ter Hofstede, A.H.M.; Barnes, J. Improving PhD Student Journeys with Process Mining: Insights from a Higher Education Institution.; de Leoni, M., Song, M., Roglinger, M., Eds.; CEUR-WS, 2021; Vol. 3112, pp. 39–49.
316. Ifenthaler, D.; Schumacher, C.; Kuzilek, J. Investigating Students' Use of Self-assessments in Higher Education Using Learning Analytics. *J Comput Assist Learn* **2023**, *39*, 255–268, doi:10.1111/jcal.12744.
317. Nicoleta, T.C. Process Mining on a Robotic Mechanism.; Institute of Electrical and Electronics Engineers Inc., 2021; pp. 205–212.
318. Lamghari, Z.; Saidi, R.; Radgui, M.; Rahmani, M.D. Mining Self-Defined Business Process in Electronic Administration. *International Journal of E-Services and Mobile Applications* **2022**, *14*, doi:10.4018/IJESMA.296572.

Prilog A Spisak primarnih studija

Industrija	Cilj analize	Poslovni proces	Naslov primarne studije	Autori	Referenca	Godina
Zdravstvo	Otkrivanje i unapređenje poslovnog procesa	Ponašanje korisnika	A Process Mining Approach to Statistical Analysis: Application to a Real-World Advanced Melanoma Dataset	Tavazzi E., i drugi	[220]	2019
		Ponašanje korisnika	Applying process mining to investigate the relation between food purchase behavior and children's weight based on the food digital cards	Matthanawongsakorn C., i drugi	[221]	2019
		Putanja pacijenta	Analyze in Depth Health Care Business Process and Key Performance Indicators using Process Mining	Elhadjamor E.A., Ghannouchi S.A.	[190]	2019
		Putanja pacijenta	Evaluating the Effectiveness of Interactive Process Discovery in Healthcare: A Case Study	Benevento E., i drugi	[191]	2019
		Putanja pacijenta	Evaluation of operational process variables in healthcare using process mining and data visualization techniques	Aguirre J.A., i drugi	[192]	2019
		Putanja pacijenta	Evaluating the process capability ratio of patients' pathways by the application of process mining, SPC and RTLS	Araghi S.N., i drugi	[193]	2019
		Proces odgovora hitne službe	Analysis of Emergency Room Episodes Duration Through Process Mining	Rojas E., i drugi	[224]	2019
		Ponašanje korisnika	Characterization of Drug Use Patterns Using Process Mining and Temporal Abstraction Digital Phenotyping	Rojas E., Capurro D.	[222]	2019

Putanja pacijenta	Modeling clinical activities based on multi-perspective declarative process mining with open EHR's characteristic	Xu H., i drugi	[194]	2020
Putanja pacijenta	Process mining project methodology in healthcare: a case study in a tertiary hospital	Pereira G.B., Santos E.A.P., Maceno M.M.C.	[30]	2020
Proces odgovora hitne službe	Mapping the patient's journey in healthcare through process mining	Arias M., i drugi	[225]	2020
Proces odgovora hitne službe	Process mining-supported emergency room process performance indicators	Cho M., i drugi	[226]	2020
Putanja pacijenta	Understanding adoption of electronic medical records: Application of process mining for health worker behavior analysis	Villamor D.A.R., Pulmano C.E., Estuar M.R.J.E.	[195]	2020
Proces odgovora hitne službe	A comparative process mining analysis of road trauma patient pathways	Andrews R., i drugi	[313]	2020
Putanja pacijenta	Supporting governance in healthcare through process mining: A case study	Agostinelli S., i drugi	[196]	2020
Putanja pacijenta	A process mining methodology for modeling unstructured processes	Stefanini A., i drugi	[197]	2020
Proces odgovora hitne službe	An ad hoc process mining approach to discover patient paths of an Emergency Department	Duma D., Aringhieri R.	[227]	2020
Putanja pacijenta	Process mining framework with time perspective for understanding acute care: a case study of AIS in hospitals	Pang J., i drugi	[198]	2021
Naplata medicinskih računa	Application of Process Mining to Medical Billing Using L* Life Cycle Model	Esiefarienrhe B.M., Omolewa I.D.	[228]	2021

Ponašanje korisnika	Seeing the signs of workarounds: A mixed-methods approach to the detection of nurses' process deviations	Beerepoot I., i drugi	[223]	2021
Tok bolesti	Process Mining of Disease Trajectories in MIMIC-III: A Case Study	Kusuma G., i drugi	[229]	2021
Putanja pacijenta	Process-aware digital twin cockpit synthesis from event logs	Bano D., i drugi	[199]	2022
Operacija	A collaborative method for simultaneous operations: case of an eye clinic	Singh S., Verma R., Koul S.	[230]	2022
Putanja pacijenta	Stable heuristic miner: Applying statistical stability to discover the common patient pathways from location event logs	Araghi S., i drugi	[200]	2022
Putanja pacijenta	A framework for the automatic description of healthcare processes in natural language: Application in an aortic stenosis integrated care process	Fontenla-Seco Y., i drugi	[201]	2022
Putanja pacijenta	Multi-perspective process mining for emergency process	Erdogan T.G., Tarhan A.K.	[202]	2022
Transport pacijenta	Evaluation of patient transport service in hospitals using process mining methods: Patients' perspective	Kropp T., Faeghi S., Lennerts K.	[231]	2022
Putanja pacijenta	OPerA: Object-Centric Performance Analysis	Park G., Adams J.N., van der Aalst W.M.P.	[203]	2022
Putanja pacijenta	Analyzing Medical Data with Process Mining: A COVID-19 Case Study	Pegoraro M., i drugi	[204]	2022
Laboratorijsko testiranje	Characterization of Laboratory Flow and Performance for Process Improvements via Application of Process Mining	Tsai E.R., i drugi	[232]	2023

	Putanja pacijenta	Analysis the patients' careflows using process mining	Rashed A.-H.M., i drugi	[205]	2023
	Prodaja	Process mining approach to performance analysis and bottleneck finding in electronic processes (case study: the billing process of hospital services)	Mohammadi F., Kazempourian S., Vanani I.R.	[233]	2023
	Nadgledanje krvnog pritiska	Extending Process Discovery with Model Complexity Optimization and Cyclic States Identification: Application to Healthcare Processes	Elkhovskaya L.O., i drugi	[234]	2023
	Putanja pacijenta	A Data-Driven Approach to Support the Understanding and Improvement of Patients' Journeys: A Case Study Using Electronic Health Records of an Emergency Department	Rismanchian F., i drugi	[206]	2023
	Poređenje poslovnih procesa	Comparing Process Models for Patient Populations: Application in Breast Cancer Care	Marazza F., i drugi	[207]	2019
		Unraveling and improving the interorganizational arthrosis care process at Maastricht UMC+: An illustration of an innovative, combined application of data and process mining	Canjels K.F., i drugi	[208]	2019
		Automatic process comparison for subpopulations: Application in cancer care	Marazza F., i drugi	[209]	2020
		Applying process mining in health technology assessment	Dallagassa M.R., i drugi	[210]	2022

	Analiza vremena čekanja	Putanja pacijenta	An Evidence-Based Decision Support Framework for Clinician Medical Scheduling	Cho M., i drugi	[211]	2019
		Putanja pacijenta	Process mining based on patient waiting time: an application in health processes	Dogan O.	[212]	2022
	Predviđanje ponašanja procesa	Putanja pacijenta	Clustering-Based Predictive Process Monitoring	Francescomarino C.D., i drugi	[213]	2019
		Putanja pacijenta	A Process Mining Application for the Analysis of Hospital-at-Home Admissions	Amantea I.A., i drugi	[214]	2020
		Putanja pacijenta	Towards Evidence-Based Analysis of Palliative Treatments for Stomach and Esophageal Cancer Patients: a Process Mining Approach	Pijnenborg P., i drugi	[215]	2021
		Putanja pacijenta	Towards the application of process mining for supporting the home hospitalization service	Aringhieri R., i drugi	[216]	2021
	Analiza promena procesa	Putanja pacijenta	Using a multi-level process comparison for process change analysis in cancer pathways	Kurniati A.P., i drugi	[217]	2020
	Analiza resursa	Putanja pacijenta	A data-driven methodology for supporting resource planning of health services	Stefanini A., i drugi	[218]	2020
		Proces odgovora hitne službe	Mining emergency event logs to support resource allocation	Li H., i drugi	[219]	2021
	Proizvodnja	Prodaja	Getting insights to improve business processes with agility: A case study using process mining	Garcia C.D.S., i drugi	[101]	2019
		Proizvodni proces	Manufacturing process improvement of offshore plant: Process mining technique and case study	Shin S.-C., i drugi	[235]	2019

Upravljanje nepredviđenim okolnostima	A business process analysis methodology based on process mining for complaint handling service processes	Wu Q., i drugi	[245]	2019
Upravljanje nepredviđenim okolnostima	Process mining in production systems	Abonyi J., Dorgo G.	[246]	2019
Nabavka	Compliance and Performance Analysis of Procurement Processes Using Process Mining BPI Challenge 2019	Diba K., Remy S., Pufahl L.	[85]	2019
Prodaja	Data-based description of process performance in end-to-end order processing	Schuh G., i drugi	[90]	2020
Proizvodni proces	Using Process Mining Approach for Machining Operations	Altan Z., Birgün S.	[236]	2020
Upravljanje nepredviđenim okolnostima	Anomaly Detection and Fault Prediction of Breakdown to Repair Process Using Mining Techniques	Bhogal R., Garg A.	[95]	2020
Planiranje proizvodnje	Integrated framework of process mining and simulation–optimization for pod structured clinical layout design	Halawa F., i drugi	[248]	2021
Proizvodni proces	A Real-World Application of Process Mining for Data-Driven Analysis of Multi-Level Interlinked Manufacturing Processes	Birk A., i drugi	[237]	2021
Prodaja	Process Mining: The first successful Peruvian case	Calderón-Ruiz G., Luque D.F.	[244]	2022
Upravljanje nepredviđenim okolnostima	Simultaneous Process Mining of Process Events and Operator Actions for Alarm Management	Bántay L., i drugi	[247]	2022

	Topologija fabrike	A Framework for Plant Topology Extraction Using Process Mining and Alarm Data	Neshastegaran A., Norouzifar A., Izadi I.	[102]	2022
	Nabavka	Object-centric process predictive analytics	Galanti R., i drugi	[242]	2023
	Nabavka	Exploiting a combined process mining approach to enhance the discovery and analysis of support processes in manufacturing	Lugaresi G., i drugi	[243]	2023
Pronalazak anomalija i nepredviđenih okolnosti	Proizvodni proces	Using Process Mining in Real-Time to Reduce the Number of Faulty Products	Nagy Z., Werner-Stark A., Dulai T.	[71]	2019
	Upravljanje promenama	Enterprise Architecture Method for Continuous Improvement of PLM Based on Process Mining	Rigger E., i drugi	[251]	2020
Analiza promena procesa	Proizvodni proces	Detecting and identifying data drifts in process event streams based on process histories	Stertz F., Rinderle-Ma S.	[70]	2019
Predviđanje ponašanja procesa	Proizvodni proces	Implementing self-* autonomic properties in self-coordinated manufacturing processes for the Industry 4.0 context	Sánchez M., Exposito E., Aguilar J.	[96]	2020
	Proizvodni proces	An extended model for remaining time prediction in manufacturing systems using process mining	Choueiri A.C., i drugi	[97]	2020
	Proizvodni proces	Performance analysis and time prediction in manufacturing systems	Ruschel E., Rocha Loures E.D.F., Santos E.A.P.	[238]	2021
Analiza vremena čekanja	Proizvodni proces	Data-Driven Analysis of Batch Processing Inefficiencies in Business Processes	Lashkevich K., i drugi	[239]	2022

		Proizvodni proces	Bottleneck Detection Through Data Integration, Process Mining and Factory Physics-Based Analytics	Kumbhar M., Ng A.H.C., Bandaru S.	[240]	2022
		Montaža	Screening Process Mining and Value Stream Techniques on Industrial Manufacturing Processes: Process Modelling and Bottleneck Analysis	Rudnitskaia J., i drugi	[249]	2022
	Otkrivanje poslovnih pravila	Proizvodni proces	Decision Mining with Time Series Data Based on Automatic Feature Generation	Scheibel B. i Rinderle-Ma S.	[241]	2022
	Automatizacija poslovnih procesa	Nabavka	Improving Purchase to Pay Process Efficiency with RPA using Fuzzy Algorithm in Process Mining	Bahaweres R.B., Amna H., Nurnaningsih D.	[250]	2022
Obrazovanje	Otkrivanje i unapređenje poslovnog procesa	Proces učenja	Implementation of Alpha Miner Algorithm in Process Mining Application Development for Online Learning Activities Based on MOODLE Event Log Data	Nafasa P., i drugi	[252]	2019
		Proces učenja	Using Process Mining Techniques to Discover Student's Activities, Navigation Paths, and Behavior in LMS Moodle	Dolak R.	[253]	2019
		Ponašanje korisnika	Describing educational trajectories of engineering students in individual high-failure rate courses that lead to late dropout	Salazar-Fernandez J.P., Sepúlveda M., Munoz-Gama J.	[271]	2019
		Proces učenja	Using Process Mining to examine the sustainability of instructional support: How stable are the effects of metacognitive prompting on self-regulatory behavior?	Sonnenberg C., Bannert M.	[254]	2019

Upravljanje resursima	Analysis of Materials and Supplies Withdrawal Process in Educational Institutions with Fuzzy Miner Technique	Meengoen P., i drugi	[270]	2020
Proces učenja	Educational Process Mining for Verifying Student Learning Paths in an Introductory Programming Course	Real E.M., i drugi	[255]	2020
Proces učenja	Data-driven Exploration of Engagement with Workplace-based Assessment in the Clinical Skills Domain	Piotrkowicz A., i drugi	[256]	2021
Proces učenja	Analysis of Student Learning Pattern in Learning Management System (LMS) using Heuristic Mining a Process Mining Approach	Chanifah S., Andreswari R., Fauzi R.	[314]	2021
Ocenjivanje studenata	Theory-based learning analytics to explore student engagement patterns in a peer review activity	Er E., i drugi	[265]	2021
Putanja studenta	Improving PhD Student Journeys with Process Mining: Insights from a Higher Education Institution	Goel K., i drugi	[315]	2021
Proces izrade softvera	Application of Constructive Modeling and Process Mining Approaches to the Study of Source Code Development in Software Engineering Courses	Shynkarenko V., Zhevaho O.	[269]	2021
Grupno rešavanje problema	Examining student regulation of collaborative, computational, problem-solving processes in openended learning environments	Emara M., i drugi	[268]	2021
Proces učenja	Using process mining for learning resource recommendation: A Moodle case study	Hachicha W., i drugi	[257]	2021

	Proces učenja	Educational process intelligence: A process mining approach and model analysis	Okoye K., Hosseini S.	[258]	2021
	Priprema nastave	Analysis and Exploratory of Lecture Preparation Process to Improve the Conformance using Process Mining	Rahmawati R., Andreswari R., Fauzi R.	[267]	2022
	Proces učenja	An Education Process Mining Framework: Unveiling Meaningful Information for Understanding Students' Learning Behavior and Improving Teaching Quality	Alqaheri H., Panda M.	[259]	2022
	Upravljanje projektima	Process Mining Methodology for Digital Processes under Smart Campus Concept	Doğan O. , Cengiz Tırpan E.	[266]	2022
	Ocenjivanje studenata	Investigating students' use of self-assessments in higher education using learning analytics	Ifenthaler D., Schumacher C., Kuzilek J.	[316]	2023
Predviđanje ponašanja procesa	Proces učenja	Generation of agent simulation models by using process mining methods on the example of E-learning process	Dorrer M., Dorrer A.	[260]	2019
Otkrivanje uzroka problema	Proces učenja	Influence of student diversity on educational trajectories in engineering high-failure rate courses that lead to late dropout	Salazar-Fernandez J.P., Sepúlveda M., Munoz-Gama J.	[261]	2019
Pronalazak anomalija i nepredviđenih aktivnosti	Upravljanje resursima	Application of Process Mining technology to analyze the correspondence of life cycles of information resources to a generalized model	Dorrer M., Popov A.	[263]	2020
Automatizacija poslovnog procesa	Ponašanje korisnika	Discovering data transfer routines from user interaction logs	Leno V., i drugi	[262]	2022
Poređenje poslovnih procesa	Ocenjivanje studenta	Analytics of Open-Book Exams with Interaction Traces in a Humanities Course	Majumdar R., i drugi	[264]	2021

Informaciono-komunikacione tehnologije	Otkrivanje i unapređenje poslovnog procesa	Proces popravke greške pri radu softvera	Monitoring the software bug-fixing process through the process mining approach	Özdağoglu G., Kavuncubaşı E.	[285]	2019
		Ponašanje korisnika	Process mining of logged gaming behavior	Ramadan S., i drugi	[277]	2019
		Proces izrade softvera	Measuring data-centre workflows complexity through process mining: the Google cluster case	Fernández-Cerero D., i drugi	[272]	2019
		Proces sajber napada	A Deep Neural Network Approach to Tracing Paths in Cybersecurity Investigations	Daniel C., i drugi	[171]	2020
		Kreiranje filogeneze virusa	Malware Phylogeny Analysis using Data-Aware Declarative Process Mining	Ardimento P., Bernardi M.L., Cimitile M.	[286]	2020
		Upravljanje nepredviđenim okolnostima	Analyzing Comments in Ticket Resolution to Capture Underlying Process Interactions	Gupta M., i drugi	[287]	2020
		Proces sajber napada	A Process Mining-based approach for Attacker Profiling	Rodriguez M., Betarte G., Calejari D.	[283]	2021
		Ponašanje korisnika	User-Centred Application for Modeling Journeys in Digital Libraries	Trabelsi M., i drugi	[278]	2021
		Ponašanje korisnika	Process Mining on Blockchain Data: A Case Study of Augur	Hobeck R., i drugi	[279]	2021
		Konfiguracija sistema	Discovering configuration workflows from existing logs using process mining	Ramos-Gutiérrez B., i drugi	[288]	2021
		Ponašanje korisnika	SOWCompact: A federated process mining method for social workflows	Rojo J., i drugi	[280]	2022

		Proces izrade softvera	Identifying Possible Improvements of Software Development Life Cycle (SDLC) Process of a Bank by Using Process Mining	Taskesenlioglu S., Ozkan N., Erdogan T.G.	[273]	2022
	Pronalazak anomalija i nepredviđenih aktivnosti	Proces izrade softvera	Monitoring a CI/CD Workflow Using Process Mining	Nogueira A.F., Zenha-Rela M.	[274]	2021
		Proizvodni proces	Process mining on a robotic mechanism	Nicoleta T.C.	[317]	2021
		Proces sajber napada	Exploring the Application of Process Mining Techniques to Improve Web Application Security	Bruno M., i drugi	[284]	2021
		Ponašanje korisnika	Scenarios for Process-Aware Insider Attack Detection in Manufacturing	Macak M., i drugi	[282]	2022
		Proces izrade softvera	Modeling and Simulating a Process Mining-Influenced Load-Balancer for the Hybrid Cloud	Azumah K., i drugi	[275]	2022
		Proces izrade softvera	Monitoring of Microservices Architecture based Applications using Process Mining	Raj V., Chander G.P.	[276]	2022
	Predviđanje ponašanja procesa	Ponašanje korisnika	Comparative Assessment of Process Mining for Supporting IoT Predictive Security	Hemmer A., i drugi	[281]	2021
Javni servisi i usluge	Otkrivanje i unapređenje poslovnog procesa	Ponašanje korisnika	Optimizing customer journey using process mining and sequence-aware recommendation	Terragni A., Hassani M.	[291]	2021
		Servisiranje automobila	Application of process mining on the example of an authorized passenger car service station in Poland	Bitkowska A., i drugi	[294]	2020
		Ponašanje korisnika	Process mining application for analysis of customer's different visits in a shopping mall	Dogan O., Fernandez-Llatas C., Oztaysi B.	[292]	2020

		Obrada tužbe	Process mining-enabled jurimetrics	Unger A.J., i drugi	[295]	2021
		Ponašanje korisnika	Estimation of Customer Activity Patterns in Open Malls by Means of Combining Localization and Process Mining Techniques	Ocana M., i drugi	[293]	2021
		Ponašanje korisnika	Mining Self-Defined Business Process in Electronic Administration	Lamghari Z., i drugi	[318]	2022
	Pronalazak anomalija i nepredviđenih aktivnosti	Proces prijave	Evaluation of Compliance Requirements for collaborative business process with process mining and a model of generic compliance controls	Gonzales L., i drugi	[289]	2022
	Otkrivanje uzroka problema	Proces prijave	Root Cause Analysis in Process Mining with Probabilistic Temporal Logic	Van Houdt G., Depaire B., Martin N.	[290]	2022
Računovodstvo, kontroling, bankarstvo i osiguranje	Pronalazak anomalija i nepredviđenih aktivnosti	Pregled godišnjeg bilansa računa	A practical process mining approach for compliance management	Becker M., Buchkremer R.	[300]	2019
	Automatizacija poslovnog procesa	Interna kontrola	Identifying the absence of effective internal controls: An alternative approach for internal control audits	Werner M., Gehrke N.	[170]	2019
	Poređenje poslovnih procesa	Obrada žalbe	Extracting Best-Practice Using Mixed-Methods: Insights and Recommendations from a Case Study in Insurance Claims Processing	Poppe E., i drugi	[299]	2021
		Prijava za kredit	TDABC and Estimation of Time Drivers Using Process Mining	Halaška M., Šperka R.	[296]	2021

	Otkrivanje i unapređenje poslovnog procesa	Obrada žalbe	Time and structural anomalies detection in business processes using process mining	Saeedi E., Esfahani F.S.	[297]	2021
		Kontroling	A case study on the application of process mining in combination with journal entry tests for financial auditing	Stephan S., Lahann J., Fettke P.	[301]	2021
	Predviđanje ponašanja procesa	Prijava za kredit	The performance assessment framework (PPAFR) for RPA implementation in a loan application process using process mining	Halaška M., Šperka R.	[298]	2022
Logistika	Analiza promena procesa	Kretanje robe	Trace Clustering exploration for detecting sudden drift: A case study in logistic process	Prathama F., i drugi	[306]	2019
	Predviđanje ponašanja procesa	Proces preuzimanja prtljaga	Predictive Process Monitoring in Operational Logistics: A Case Study in Aviation	Gunnarsson B.R., vanden Broucke S.K.L.M., De Weerd J.	[305]	2019
	Analiza resursa	Montaža	Enabling value stream mapping for internal logistics using multidimensional process mining	Knoll D., Reinhart G., Prüglermeier M.	[88]	2019
	Otkrivanje i unapređenje poslovnog procesa	Upravljanje skladištem	Analysis of Warehouse Operations using Process Mining Techniques: A Case Study of Scientific Laboratory Equipment and Facilities	Montianrat S., i drugi	[304]	2020
		Nabavka	Analysis of E-Commerce Process in the Downstream Section of Supply Chain Management Based on Process and Data Mining	Tridalestari F.A., i drugi	[302]	2022
		Prodaja	Data-driven performance management of business units using process mining and DEA: case study of an Iranian chain store	Maddah N., Roghanian E.	[303]	2023

Građevina	Predviđanje ponašanja procesa	Izgradnja	Integrating Process Mining with Discrete-Event Simulation for Dynamic Productivity Estimation in Heavy Civil Construction Operations	Rashid K.M., Louis J.	[310]	2022
	Otkrivanje i unapređenje poslovnog procesa	Izgradnja	Data analytic approaches for mining process improvement-machinery utilization use case	Brzychczy E., Gackowiec P., Liebetrau M.	[307]	2020
		Izgradnja	Process Discovery and Conformance Checking in Modular Construction Using RFID and Process Mining	Rashid K.M., Louis J.	[308]	2020
		Izgradnja	Automated process discovery from event logs in BIM construction projects	Pan Y., Zhang L.	[309]	2021
		Procena troškova	An Approach to Data Driven Process Discovery in the Cost Estimation Process of a Construction Company	Kropp T., Bombeck A., Lennerts K.	[311]	2021
Energija	Otkrivanje i unapređenje poslovnog procesa	Upravljanje bezbednosnim rizikom	Using Big Data to Improve Safety Performance: An Application of Process Mining to Enhance Data Visualisation	Pika A., i drugi	[312]	2021

Prilog B Upitnik

Kvalitet podataka u automatskom otkrivanju poslovnih procesa:

Problemi i rešenja

1. Uvod

Kao što je već dobro poznato u zajednici automatskog otkrivanja poslovnih procesa, kvalitet podataka u dnevniku izvršenja događaja ima ogroman uticaj na rezultate analize. Cilj ovog upitnika je da objedini uvid praktikanata i istraživača u probleme kvaliteta podataka i tehnike čišćenja dnevnika izvršenja događaja, i način na koji se problemi kvaliteta zapravo uklanjaju u praksi. Takođe, upitnik prikuplja i sumira demografske informacije o praktikantima i istraživačim iz oblasti širom sveta.

Upitnik je podeljen na 4 stranice, sa sledećim temama:

1. Demografija;
2. Problemi kvaliteta podataka dnevnika izvršenja događaja;
3. Tehnika čišćenja dnevnika izvršenja događaja;
4. Odabir tehnika čišćenja dnevnika izvršenja događaja .

Pitanja su usmerena na praktikantima i istraživačima automatskog otkrivanja poslovnih procesa sa iskustvom. Odgovori su anonimni i analizirani su za potrebe izrade doktorske disertacije na Fakultetu tehničkih nauka na Univerzitetu u Novom Sadu, Srbija.

Hvala na učešću u istraživanju!

2. Demografija

Ova sekcija sadrži osnovna pitanja o ispitanicima i njihovom poznavanju oblasti. Molim odgovorite na sva pitanja.

1. Kako biste ocenili vaše poznavanje oblasti obrade podataka, uopšteno?

- Slabo
- Umereno dobro
- Dobro
- Veoma dobro
- Odlično

2. Koja je vaša uloga u zajednici automatskog otkrivanja poslovnih procesa?

- Istraživač
- Praktikant
- Oba
- Drugo (molim specificirati)

3. Koliko dugo istražujete i/ili primenjujete automatsko otkrivanje poslovnih procesa?

- Manje od 1 godine
- 1 - 5 godina
- 5 - 10 godina
- Više od 10 godina

4. Molim navedite vaše zanimanje.

5. U kojoj zemlji trenutno radite?

6. Koji softver koristite najviše prilikom izvođenja automatskog otkrivanja poslovnih procesa? Odaberite jedan ili dodajte ukoliko nije naveden.

- *ProM*
- *Celonis*

- *Fluxicon Disco*
- *RapidProm*
- *IBM Process Mining*
- *SAP Signavio Business Intelligence*
- *ARIS Process Mining*
- *MPM ProcessMining*
- *QPR ProcessAnalyzer*
- *Apromore*
- *Apian Process Mining*
- Drugi (Molim navesti)

7. Koji softver koristite najviše prilikom izvođenja čišćenja dnevnika izvršenja događaja? Odaberite jedan ili dodajte ukoliko nije naveden.

- *ProM*
- *Celonis*
- *Fluxicon Disco*
- *RapidProm*
- *IBM Process Mining*
- *SAP Signavio Business Intelligence*
- *ARIS Process Mining*
- *MPM ProcessMining*
- *QPR ProcessAnalyzer*
- *Apromore*
- *Apian Process Mining*
- Drugi (Molim navesti)

3. Problemi kvaliteta podataka

Ova sekcija upitnika sadrži taksonomiju problema kvaliteta podataka dnevnika izvršenja događaja određenu kategorijama problema kvaliteta podataka:

1. Nedostajući podaci (podaci ne postoje u dnevniku podataka),
2. Netačni podaci (podaci postoje, ali su pogrešno zabeleženi),
3. Neprecizni podaci (podaci su previše neodređeni, dovodeći to gubitka preciznosti), i
4. Irelevantni podaci (podaci nisu od značaja za analizu),

koji se manifestuju kroz entitete dnevnika izvršenja događaja (događaj, slučaj, naziv aktivnosti, itd.).

Radi lakšeg razumevanja svakog problema kvaliteta, dodato je kratko objašnjenje i naziv obrasca imperfekcije, ukoliko postoji u određenom slučaju.

Molim odgovorite na svako pitanje vezano za značaj i učestalost pojave navedenih problema kvaliteta.

10. Prema Vašem iskustvu, koliko su značajni sledeći problemi kvaliteta?

	Nije značajan	Malo značajan	Umereno značajan	Značajan	Veoma značajan
Nedostajući podaci: Slučaj					
Ovaj problem kvaliteta se javlja u scenariju kada je neki slučaj procesa izvršen u realnosti, a nije					

zabeležen u dnevniku izvršenja događaja.
Nedostajući podaci: Događaj (Raštrkani događaj)
Ovaj problem kvaliteta se javlja u scenariju kada jedan ili više događaja nedostaje unutar putanje procesa, iako su se izvršili u stvarnosti.
Nedostajući podaci: Veza (Neuhvatljiv slučaj)
Ovaj problem kvaliteta se javlja u scenariju kada nije zabeležena veza između slučaja i događaja.
Nedostajući podaci: Naziv aktivnosti
Ovaj problem kvaliteta se javlja u scenariju kada nazivi aktivnosti događaja nisu zabeleženi.
Nedostajući podaci: Atribut slučaja i/ili događaja
Ovaj problem kvaliteta se javlja u scenariju kada nazivi atributa slučaja i/ili događaja nisu zabeleženi.
Nedostajući podaci: Vremenska odrednica
Ovaj problem kvaliteta se javlja u scenariju kada vremenska odrednica ne postoji kod jednog ili više događaja.
Nedostajući podaci: Resurs
Ovaj problem kvaliteta se javlja u scenariju kada nisu zabeleženi resursi koji su utrošeni na izvođenju aktivnosti.
Netačni podaci: Slučaj
Ovaj problem kvaliteta se javlja u scenariju kada određeni slučajevi procesa pripadaju pogrešnom dnevniku izvršenja događaja.
Netačni podaci: Događaj
Ovaj problem kvaliteta se javlja u scenariju kada su određeni

dogadjaji u dnevniku izvršenja dogadjaja pogrešno zabeleženi.
Netačni podaci: Veza (Raštrkani slučaj)
Ovaj problem kvaliteta se javlja kada su veze između dogadjaja i slučajeva pogrešno zabeležene.
Netačni podaci: Naziv aktivnosti (Zagađeni naziv atributa, Iskrivljeni naziv atributa)
Ovaj problem kvaliteta podataka se javlja kada su nazivi aktivnosti netačno zabeleženi.
Netačni podaci: Atribut slučaja i/ili dogadjaja
Ovaj problem kvaliteta se javlja u scenariju kada su nazivi atributa slučaja i/ili dogadjaja netačno zabeleženi.
Netačni podaci: Vremenska odrednica (Hvatanje dogadjaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj)
Ovaj problem kvaliteta odataka se javlja kada zabeležene vremenske odrednice ne odgovaraju stvarnom vremenu u kojem su se događaji izvršili.
Netačni podaci: Resurs (Zagađeni naziv atributa)
Ovaj problem kvaliteta se javlja u scenariju kada su zabeleženi resursi koji su utrošeni na izvođenju aktivnosti netačni.
Neprecizni podaci: Veza
Ovaj problem kvaliteta se odnosi na scenario kada zbog odabrane definicije slučaja, neki događaji ne mogu da se povežu sa drugim tipom slučaja.
Neprecizni podaci: Naziv aktivnosti (Istoimeni naziv aktivnosti)
Ovaj problem kvaliteta odgovara scenariju kada su nazivi aktivnosti

previše uopšteni. Kao posledica, unutar jedan putanje procesa može biti više aktivnosti sa istim nazivom, a nisu iste aktivnosti.					
Neprecizni podaci: Atribut slučaja i/ili događaja (Sinonimni naziv atributa)					
Ovaj problem kvaliteta se javlja kada su nazivi atributa slučaja i/ili događaja previše uopšteni.					
Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj)					
Ovaj problem kvaliteta se javlja kada je vremenska odrednica previše uopštena, na previše visokom nivou abstrakcije da bi bila upotrebljiva.					
Neprecizni podaci: Resurs					
Ovaj problem kvaliteta se javlja u scenariju kada su zabeleženi resursi koji su utrošeni na izvođenju aktivnosti previše uopšteni, iako u realnom sistemu postoje konkretni nazivi resursa.					
Irelevantni podaci: Slučaj					
Ovaj problem kvaliteta se javlja u scenariju kada su zabeleženi podaci o slučajevima neznčajni za kontekst analize.					
Irelevantni podaci: Događaj (Hvatanje događaja zasnovano na obrascu, Kolateralni događaji)					
Ovaj problem kvaliteta se javlja u scenariju kada su zabeleženi podaci o događajima neznčajni za kontekst analize.					
Količina podataka, nivo detaljnosti događaja, kompleksnost					
11. Molim označite koliko često se susrećete sa sledećim problemima kvaliteta.					
	Nikad	Retko	Povremeno	Često	Veoma često
Nedostajući podaci: Slučaj					
Ovaj problem kvaliteta se javlja u scenariju kada je neki slučaj procesa izvršen u realnosti, a nije					

zabeležen u dnevniku izvršenja događaja.
Nedostajući podaci: Događaj (Raštrkani događaj)
Ovaj problem kvaliteta se javlja u scenariju kada jedan ili više događaja nedostaje unutar putanje procesa, iako su se izvršili u stvarnosti.
Nedostajući podaci: Veza (Neuhvatljiv slučaj)
Ovaj problem kvaliteta se javlja u scenariju kada nije zabeležena veza između slučaja i događaja.
Nedostajući podaci: Naziv aktivnosti
Ovaj problem kvaliteta se javlja u scenariju kada nazivi aktivnosti događaja nisu zabeleženi.
Nedostajući podaci: Atribut slučaja i/ili događaja
Ovaj problem kvaliteta se javlja u scenariju kada nazivi atributa slučaja i/ili događaja nisu zabeleženi.
Nedostajući podaci: Vremenska odrednica
Ovaj problem kvaliteta se javlja u scenariju kada vremenska odrednica ne postoji kod jednog ili više događaja.
Nedostajući podaci: Resurs
Ovaj problem kvaliteta se javlja u scenariju kada nisu zabeleženi resursi koji su utrošeni na izvođenju aktivnosti.
Netačni podaci: Slučaj
Ovaj problem kvaliteta se javlja u scenariju kada određeni slučajevi procesa pripadaju pogrešnom dnevniku izvršenja događaja.
Netačni podaci: Događaj
Ovaj problem kvaliteta se javlja u scenariju kada su određeni

dogadjaji u dnevniku izvršenja dogadjaja pogrešno zabeleženi.
Netačni podaci: Veza (Raštrkani slučaj)
Ovaj problem kvaliteta se javlja kada su veze između događaja i slučajeva pogrešno zabeležene.
Netačni podaci: Naziv aktivnosti (Zagađeni naziv atributa, Iskrivljeni naziv atributa)
Ovaj problem kvaliteta podataka se javlja kada su nazivi aktivnosti netačno zabeleženi.
Netačni podaci: Atribut slučaja i/ili događaja
Ovaj problem kvaliteta se javlja u scenariju kada su nazivi atributa slučaja i/ili događaja netačno zabeleženi.
Netačni podaci: Vremenska odrednica (Hvatanje događaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj)
Ovaj problem kvaliteta odataka se javlja kada zabeležene vremenske odrednice ne odgovaraju stvarnom vremenu u kojem su se događaji izvršili.
Netačni podaci: Resurs (Zagađeni naziv atributa)
Ovaj problem kvaliteta se javlja u scenariju kada su zabeleženi resursi koji su utrošeni na izvođenju aktivnosti netačni.
Neprecizni podaci: Veza
Ovaj problem kvaliteta se odnosi na scenario kada zbog odabrane definicije slučaja, neki događaji ne mogu da se povežu sa drugim tipom slučaja.
Neprecizni podaci: Naziv aktivnosti (Istoimeni naziv aktivnosti)
Ovaj problem kvaliteta odgovara scenariju kada su nazivi

aktivnosti previše uopšteni. Kao posledica, unutar jedan putanje procesa može biti više aktivnosti sa istim nazivom, a nisu iste aktivnosti.
Neprecizni podaci: Atribut slučaja i/ili događaja (Sinonimni naziv atributa)
Ovaj problem kvaliteta se javlja kada su nazivi atributa slučaja i/ili događaja previše uopšteni.
Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj)
Ovaj problem kvaliteta se javlja kada je vremenska odrednica previše uopštena, na previše visokom nivou abstrakcije da bi bila upotrebljiva.
Neprecizni podaci: Resurs
Ovaj problem kvaliteta se javlja u scenariju kada su zabeleženi resursi koji su utrošeni na izvođenju aktivnosti previše uopšteni, iako u realnom sistemu postoje konkretni nazivi resursa.
Irelevantni podaci: Slučaj
Ovaj problem kvaliteta se javlja u scenariju kada su zabeleženi podaci o slučajevima neznčajni za kontekst analize.
Irelevantni podaci: Događaj (Hvatanje događaja zasnovano na obrascu, Kolateralni događaji)
Ovaj problem kvaliteta se javlja u scenariju kada su zabeleženi podaci o događajima neznčajni za kontekst analize.
Količina podataka, nivo detaljnosti događaja, kompleksnost

4. Tehnike čišćenja dnevnika izvršenja događaja

Ova sekcija upitnika ima za cilj prikupljanje informacija o različitim tehnikama korišćenim radi pripreme dnevnika izvršenja događaja za analizu. Pitanja su fokusirana na percipirani značaj i učestalost primene tehnika i sadrže tehnike grupisane na osnovu pregleda literature:

1. Klasterovanje putanja (npr., Dodatak za klasterovanje putanja u ProM alatu, minimalno drvo raspona klasterovanje, klasterovanje zasnovano na statističkoj inferentnosti, klasterovanje putanja na osnovu K-srednje vrednosti),
2. Tehnike popravke dnevnika izvršenja događaja (npr., Heuristički dodatak za popravku u ProM alatu, dodatak za popravku dnevnika u ProM alatu),
3. Filtriranje putanje/događaja (npr., Filter retkog ponašanja, filtriranje aktivnosti zasnovano na entropiji, algoritam grananja i spajanja),
4. Abstrakcija događaja (npr., Semantička abstrakcija),
5. Mašinsko učenje, veštačka inteligencija i duboko učenje (npr., Bajezove mreže, algoritam grananja i spajanja, LSTM veštačke neuronske mreže, algoritam stabla odlučivanja - CART),
6. Tehnike poravnavanja (npr., Poravnanje zasnovano na trošku, proveru usaglašenosti zasnovana na poravnavanju, poravnavanje putanja, dodatak za podudaranje putanja u ProM alatu),
7. Ugrađena obrada podataka (Induktivni algoritam za otkrivanje procesa, *Split miner* algoritam za otkrivanje procesa, *ILP miner* algoritam za otkrivanje procesa i interaktivni pristup otkrivanja modela procesa).

12. Prema Vašem iskustvu, koliko su značajne sledeće grupe tehnika čišćenja dnevnika izvršenja događaja, kako bi se smanjili ili uklonili problemi kvaliteta podataka?

	Nisu značajni	Malo značajni	Umereno značajni	Značajni	Veoma značajni
Klasterovanje putanja					
Tehnike popravke dnevnika izvršenja događaja					
Filtriranje putanje/događaja					
Abstrakcija događaja					
Mašinsko učenje, veštačka inteligencija i duboko učenje					
Tehnike poravnavanja					
Ugrađena obrada podataka					
Drugo (Molim navesti)					

13. Molim Vas odaberite koliko često koristite navedene grupe tehnika čišćenja dnevnika izvršenja događaja.

	Nikad	Retko	Povremeno	Često	Veoma često
Klasterovanje putanja					
Tehnike popravke dnevnika izvršenja događaja					

Filtriranje putanje/događaja
Abstrakcija događaja
Mašinsko učenje, veštačka inteligencija i duboko učenje
Tehnike poravnavanja
Ugrađena obrada podataka
Drugo (Molim navesti)

5. Odabir tehnika čišćenja dnevnika izvršenja događaja

U ovoj sekciji upitnika potrebno je odabrati jednu grupu tehnika čišćenja dnevnika izvršenja događaja koja je, prema vašem iskustvu, najpogodnija za primenu kod određenih problema dnevnika izvršenja događaja.

14. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Nedostajući podaci: Slučaj.

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje
- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

15. Molim odaberite jednu grupu tehnika koje biste upotreabili da umanjite ili uklonite sledeći problem kvaliteta podataka - Nedostajući podaci: Događaj (Raštrkani događaj).

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje
- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

16. Molim odaberite jednu grupu tehnika koje biste upotreabili da umanjite ili uklonite sledeći problem kvaliteta podataka - Nedostajući podaci: Veza (Neuhvatljiv slučaj).

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje
- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

17. Molim odaberite jednu grupu tehnika koje biste upotreabili da umanjite ili uklonite sledeći problem kvaliteta podataka - Netačni podaci: Naziv aktivnosti (Zagađeni naziv atributa, Iskrivljeni naziv atributa).

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje
- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

18. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Nedostajući podaci: Atribut slučaja i/ili događaja.

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje
- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

18. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Nedostajući podaci: Vremenska odrednica.

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje
- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

19. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Nedostajući podaci: Resurs.

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje
- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

20. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Netačni podaci: Slučaj.

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje
- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

21. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Netačni podaci: Događaj.

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja

- Abstrakcija događaja
 - Mašinsko učenje, veštačka inteligencija i duboko učenje
 - Tehnike poravnavanja
 - Ugrađena obrada podataka
 - Drugo (Molim navesti)
22. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Netačni podaci: Veza (Raštrkani slučaj).
- Klasterovanje putanja
 - Tehnike popravke dnevnika izvršenja događaja
 - Filtriranje putanje/događaja
 - Abstrakcija događaja
 - Mašinsko učenje, veštačka inteligencija i duboko učenje
 - Tehnike poravnavanja
 - Ugrađena obrada podataka
 - Drugo (Molim navesti)
23. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Netačni podaci: Naziv aktivnosti (Zagađeni naziv atributa, Iskrivljeni naziv atributa)
- Klasterovanje putanja
 - Tehnike popravke dnevnika izvršenja događaja
 - Filtriranje putanje/događaja
 - Abstrakcija događaja
 - Mašinsko učenje, veštačka inteligencija i duboko učenje
 - Tehnike poravnavanja
 - Ugrađena obrada podataka
 - Drugo (Molim navesti)
24. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Netačni podaci: Atribut slučaja i/ili događaja.
- Klasterovanje putanja
 - Tehnike popravke dnevnika izvršenja događaja
 - Filtriranje putanje/događaja
 - Abstrakcija događaja
 - Mašinsko učenje, veštačka inteligencija i duboko učenje
 - Tehnike poravnavanja
 - Ugrađena obrada podataka
 - Drugo (Molim navesti)
25. Molim odaberite jednu grupu tehnika koje biste upotreabili da umanjite ili uklonite sledeći problem kvaliteta podataka - Netačni podaci: Vremenska odrednica (Hvatanje događaja zasnovano na obrascu, Nehotično putovanje kroz vreme, Neukorenjeni događaj).
- Klasterovanje putanja
 - Tehnike popravke dnevnika izvršenja događaja
 - Filtriranje putanje/događaja
 - Abstrakcija događaja
 - Mašinsko učenje, veštačka inteligencija i duboko učenje
 - Tehnike poravnavanja
 - Ugrađena obrada podataka
 - Drugo (Molim navesti)
26. Molim odaberite jednu grupu tehnika koje biste upotreabili da umanjite ili uklonite sledeći problem kvaliteta podataka - Netačni podaci: Resurs (Zagađeni naziv atributa).
- Klasterovanje putanja
 - Tehnike popravke dnevnika izvršenja događaja

- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje

- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

27. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Neprecizni podaci: Veza.

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje

- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

28. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Neprecizni podaci: Naziv aktivnosti (Istoimeni naziv aktivnosti).

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje

- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

29. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Neprecizni podaci: Atribut slučaja i/ili događaja (Sinonimni naziv atributa).

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje

- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

30. Molim odaberite jednu grupu tehnika koje biste upotreabili da umanjite ili uklonite sledeći problem kvaliteta podataka - Neprecizni podaci: Vremenska odrednica (Neukorenjeni događaj).

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje

- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

31. Molim odaberite jednu grupu tehnika koje biste upotreabili da umanjite ili uklonite sledeći problem kvaliteta podataka - Neprecizni podaci: Resurs.

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje

- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

32. Molim odaberite jednu grupu tehnika koje biste upotreбили da umanjite ili uklonite sledeći problem kvaliteta podataka - Irelevantni podaci: Slučaj.

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje

- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

33. Molim odaberite jednu grupu tehnika koje biste upotreabili da umanjite ili uklonite sledeći problem kvaliteta podataka - Irelevantni podaci: Događaj (Hvatanje događaja zasnovano na obrascu, Kolateralni događaji).

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje

- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

34. Molim odaberite jednu grupu tehnika koje biste upotreabili da umanjite ili uklonite sledeći problem kvaliteta podataka - Količina podataka, nivo detaljnosti događaja, kompleksnost.

- Klasterovanje putanja
- Tehnike popravke dnevnika izvršenja događaja
- Filtriranje putanje/događaja
- Abstrakcija događaja
- Mašinsko učenje, veštačka inteligencija i duboko učenje
- Tehnike poravnavanja
- Ugrađena obrada podataka
- Drugo (Molim navesti)

35. Ukoliko ste zainteresovani za rezultate istraživanja, molim navedite Vaš e-mail.

Овај Образац чини саставни део докторске дисертације, односно докторског уметничког пројекта који се брани на Универзитету у Новом Саду. Попуњен Образац укоричити иза текста докторске дисертације, односно докторског уметничког пројекта.

План третмана података

Назив пројекта/истраживања
Кључни фактори примене аутоматског откривања пословних процеса у индустријским системима
Назив институције/институција у оквиру којих се спроводи истраживање
а) Факултет техничких наука, Нови Сад, Србија б) в)
Назив програма у оквиру ког се реализује истраживање
1. Опис података
1.1 Врста студије Укратко описати тип студије у оквиру које се подаци прикупљају <u>Студија спроведена у оквиру дисертације је обухватила квантитативно истраживање са циљем идентификовања кључних фактора примене аутоматског откривања пословних процеса у индустријским системима.</u> 1.2 Врсте података а) квантитативни б) квалитативни 1.3. Начин прикупљања података а) анкете, упитници, тестови б) клиничке процене, медицински записи, електронски здравствени записи в) генотипови: навести врсту _____ г) административни подаци: навести врсту _____ д) узорци ткива: навести врсту _____



ђ) снимци, фотографије: навести врсту _____

е) текст, навести врсту _____

ж) мапа, навести врсту _____

з) остало: систематски преглед литературе

1.3 Формат података, употребљене скале, количина података

1.3.1 Употребљени софтвер и формат датотеке:

а) **Excel фајл, датотека** .xls

б) **SPSS фајл, датотека** .sav

с) PDF фајл, датотека _____

д) Текст фајл, датотека _____

е) JPG фајл, датотека _____

ф) Остало, датотека _____

1.3.2. Број записа (код квантитативних података)

а) број варијабли **36**

б) број мерења (испитаника, процена, снимака и сл.) **202**

1.3.3. Поновљена мерења

а) да

б) не

Уколико је одговор да, одговорити на следећа питања:

а) временски размак измедју поновљених мера је _____

б) варијабле које се више пута мере односе се на _____

в) нове верзије фајлова који садрже поновљена мерења су именоване као _____

Напомене: _____

Да ли формати и софтвер омогућавају дељење и дугорочну валидност података?

а) Да

б) Не

Ако је одговор не, образложити _____

2. Прикупљање података

2.1 Методологија за прикупљање/генерисање података

Методологија прикупљања података истраживачког модела 1 јесте систематски преглед литературе студија случаја из области примене аутоматског откривања пословних процеса. Методологија прикупљања података истраживачког модела 2 јесте упитник као мерни инструмент, који је електронски дистрибуиран испитаницима.

2.1.1. У оквиру ког истраживачког нацрта су подаци прикупљени?

а) експеримент, навести тип _____

б) корелационо истраживање, навести тип: Хи-квадрат тест, Фишеров егзактни тест, Анализа варијанси, Пирсонов тест корелације

ц) анализа текста, навести тип _____

д) остало, навести шта Дескриптивна статистичка анализа података

2.1.2 Навести врсте мерних инструмената или стандарде података специфичних за одређену научну дисциплину (ако постоје).

Протокол систематског прегледа литературе

Електронски упитник формиран на основу теоријских подлога

2.2 Квалитет података и стандарди

Пречишћавање и трансформација података пре уласка у процесе анализе.

2.2.1. Третман недостајућих података

а) Да ли матрица садржи недостајуће податке? Да **Не**

Ако је одговор да, одговорити на следећа питања:

а) Колики је број недостајућих података? _____

б) Да ли се кориснику матрице препоручује замена недостајућих података? Да Не

в) Ако је одговор да, навести сугестије за третман замене недостајућих података

2.2.2. На који начин је контролисан квалитет података? Описати

Како би се обезбедила валидност резултата истраживања, елиминисани су непотпуни одговори кроз поступак иницијалног пречишћавања података. Такође, елиминисани су испитианици чије је искуство у области аутоматског откривања пословних процеса процењено као „слабо“.

2.2.3. На који начин је извршена контрола уноса података у матрицу?

Софтвер који је коришћен аутоматски генерише податке које је аутор претходно обрадио.

Након генерисања података, извршена је провера увидом у иницијалну базу података.

3. Третман података и пратећа документација

3.1. Третман и чување података

3.1.1. Подаци ће бити депоновани у Репозиторијум докторских дисертација Универзитета у Новом Саду.

3.1.2. URL адреса <https://www.cris.uns.ac.rs/searchDissertations.jsf>

3.1.3. DOI _____

3.1.4. Да ли ће подаци бити у отвореном приступу?

а) **Да**

б) Да, али после ембарга који ће трајати до _____

в) Не

Ако је одговор не, навести разлог _____

3.1.5. Подаци неће бити депоновани у репозиторијум, али ће бити чувани.

Образложење

3.2. Метаподаци и документација података

3.2.1. Који стандард за метаподатке ће бити примењен?

Стандард који примењује Репозиторијум Универзитета у Новом Саду.

3.2.1. Навести метаподатке на основу којих су подаци депоновани у репозиторијум.

Душанка Дакић (2023): Кључни фактори примене аутоматског откривања пословних процеса у индустријским системима

Ако је потребно, навести методе које се користе за преузимање података, аналитичке и процедуралне информације, њихово кодирање, детаљне описе варијабли, записа итд.

3.3 Стратегија и стандарди за чување података

3.3.1. До ког периода ће подаци бити чувани у репозиторијуму? Нема ограничења.

3.3.2. Да ли ће подаци бити депоновани под шифром? Да **Не**

3.3.3. Да ли ће шифра бити доступна одређеном кругу истраживача? Да **Не**

3.3.4. Да ли се подаци морају уклонити из отвореног приступа после извесног времена?

Да **Не**

Образложити

4. Безбедност података и заштита поверљивих информација

Овај одељак МОРА бити попуњен ако ваши подаци укључују личне податке који се односе на учеснике у истраживању. За друга истраживања треба такође размотрити заштиту и сигурност података.

4.1 Формални стандарди за сигурност информација/података

Истраживачи који спроводе испитивања с људима морају да се придржавају Закона о заштити података о личности (https://www.paragraf.rs/propisi/zakon_o_zastiti_podataka_o_licnosti.html) и одговарајућег институционалног кодекса о академском интегритету.

4.1.2. Да ли је истраживање одобрено од стране етичке комисије? Да **Не**

Ако је одговор Да, навести датум и назив етичке комисије која је одобрила истраживање

4.1.2. Да ли подаци укључују личне податке учесника у истраживању? Да **Не**

Ако је одговор да, наведите на који начин сте осигурали поверљивост и сигурност информација везаних за испитанике:

- а) Подаци нису у отвореном приступу
- б) Подаци су анонимизирани
- ц) Остало, навести шта

5. Доступност података

5.1. Подаци ће бити

а) **јавно доступни**

б) доступни само уском кругу истраживача у одређеној научној области

ц) затворени

Ако су подаци доступни само уском кругу истраживача, навести под којим условима могу да их користе:

Ако су подаци доступни само уском кругу истраживача, навести на који начин могу приступити подацима:

5.4. Навести лиценцу под којом ће прикупљени подаци бити архивирани.



6. Улоге и одговорност

6.1. Навести име и презиме и мејл адресу власника (аутора) података

Душанка Дакић, dakic.dusanka@uns.ac.rs

6.2. Навести име и презиме и мејл адресу особе која одржава матрицу с подацима

Душанка Дакић, dakic.dusanka@uns.ac.rs

6.3. Навести име и презиме и мејл адресу особе која омогућује приступ подацима другим истраживачима

Душанка Дакић, dakic.dusanka@uns.ac.rs
