

# Mogućnosti primjene alata poslovne inteligencije u maloserijskoj proizvodnji

---

**Savić, Marko**

**Master's thesis / Diplomski rad**

**2014**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture / Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:235:978368>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2025-01-31**

*Repository / Repozitorij:*

[Repository of Faculty of Mechanical Engineering and Naval Architecture University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU  
FAKULTET STROJARSTVA I BRODOGRADNJE

# **DIPLOMSKI RAD**

**Marko Savić**

Zagreb, 2014.

SVEUČILIŠTE U ZAGREBU  
FAKULTET STROJARSTVA I BRODOGRADNJE

# **DIPLOMSKI RAD**

Mentor:

Prof. dr. sc. Dragutin Lisjak

Student:

Marko Savić

Zagreb, 2014.

Izjavljujem da sam diplomski rad na temu „Mogućnosti primjerne alata poslovne inteligencije u maloserijskoj proizvodnji“ izradio samostalno koristeći stečena znanja tijekom studija i navedenu literaturu.

Ovom prilikom želio bih prije svega zahvaliti prof. dr. sc. Dragutinu Lisjaku na utrošenom trudu, vremenu i savjetima prilikom izrade ovog diplomskog rada.

Želim se zahvaliti i kolegi mag. ing. mech. Davoru Kolaru na svim savjetima i pomoći prilikom izrade rada.

Također se želim zahvaliti i svojim roditeljima, kao i svima koji su mi na bilo koji način pomogli tijekom studija.

Marko Savić



SVEUČILIŠTE U ZAGREBU  
**FAKULTET STROJARSTVA I BRODOGRADNJE**



Središnje povjerenstvo za završne i diplomske ispite  
Povjerenstvo za diplomske ispite studija strojarstva za smjerove:  
proizvodno inženjerstvo, računalno inženjerstvo, industrijsko inženjerstvo i menadžment, inženjerstvo  
materijala i mehatronika i robotika

Sveučilište u Zagrebu Fakultet strojarstva i brodogradnje	
Datum	Prilog
Klasa:	
Ur.broj:	

## DIPLOMSKI ZADATAK

Student: **MARKO SAVIĆ**

Mat. br.: 0035174247

Naslov rada na  
hrvatskom jeziku:

**MOGUĆNOSTI PRIMJENE ALATA POSLOVNE INTELIGENCIJE  
U MALOSERIJSKOJ PROIZVODNJI**

Naslov rada na  
engleskom jeziku:

**APPLICATION OF BUSINESS INTELIGENCE TOOLS IN THE  
SMALL-SERIES PRODUCTION**

Opis zadatka:

Proizvodno poduzeće djeluje u okruženju u kojem je potrebno brzo i učinkovito donošenje odluka kako bi se osigurao proizvod u skladu sa potrebama kupca. Današnji informacijski sustavi sposobni su bilježiti i prikupljati podatke iz svih procesa proizvodnje. Rezultat je velika količina generiranih podataka čije je tumačenje u cilju donošenja neke poslovne odluke otežano. To na kraju dovodi do kašnjenja cijelog procesa donošenja odluka što pak rezultira sporim odazivom prema potencijalnom kupcu. Zbog toga se u današnjim poslovnim sustavima sve više primjenjuju metode poslovne inteligencije koje omogućavaju učinkovitije tumačenje velike količine podataka prikupljenih iz procesa proizvodnje. U radu je potrebno obraditi slijedeće:

1. Opisati metode poslovne inteligencije.
2. Detaljno obraditi metodu rudarenja podataka (*data mining*).
3. Objasniti potrebu implementacije metoda poslovne inteligencije u proizvodne sustave maloserijske proizvodnje.
4. Predložiti i razraditi konkretan primjer problema iz maloserijske proizvodnje na kojem će se dokazati da primjena metoda poslovne inteligencije u konačnici rezultira bržim i učinkovitijim poslovnim odlukama.

Zadatak zadan:

Rok predaje rada:

Predviđeni datum obrane:


25. rujna 2014.

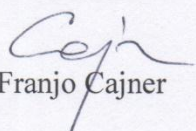
27. studenog 2014.

3., 4. i 5. prosinca 2014.

Zadatak zadao:

Predsjednik Povjerenstva:

  
Prof.dr.sc. Dragutin Lisjak

  
Prof. dr. sc. Franjo Čajner

## SADRŽAJ

<b>SADRŽAJ .....</b>	<b>I</b>
<b>POPIS SLIKA.....</b>	<b>IV</b>
<b>POPIS TABLICA .....</b>	<b>VI</b>
<b>POPIS KRATICA .....</b>	<b>VII</b>
<b>SAŽETAK .....</b>	<b>VIII</b>
<b>SUMMARY .....</b>	<b>IX</b>
<b>1. UVOD .....</b>	<b>1</b>
<b>2. UVOD U POSLOVNU INTELIGENCIJU .....</b>	<b>2</b>
2.1. Značaj učinkovitih i pravovremenih odluka .....	2
2.1.1. Učinkovite odluke .....	3
2.1.2. Pravovremene odluke .....	3
2.2. Podaci, informacije i znanje .....	4
2.2.1. Podaci .....	5
2.2.2. Informacije.....	5
2.2.3. Znanje .....	5
2.3. Uloga matematičkih modela.....	6
2.4. Arhitektura sustava poslovne inteligencije .....	7
2.4.1. Izvori podataka (eng. Data sources).....	7
2.4.2. Skladišta podataka (eng. Data warehouses and Data marts) .....	7
2.4.3. Metodologije poslovne inteligencije .....	7
2.4.4. Istraživanje podataka.....	9
2.4.5. Rudarenje podataka.....	9
2.4.6. Optimizacija.....	10
2.4.7. Odluka .....	10
2.5. Ciklus analize sustava poslovne inteligencije .....	10
2.5.1. Analiza.....	11
2.5.2. Uvid.....	11
2.5.3. Odluka .....	11
2.5.4. Evaluacija .....	11
2.6. Ključni faktori razvoja sustava poslovne inteligencije .....	11
2.6.1. Tehnologija .....	12
2.6.2. Analiza.....	12
2.6.3. Ljudski potencijali.....	12
2.7. Razvoj sustava poslovne inteligencije .....	13
2.7.1. Analiza.....	14
2.7.2. Dizajn .....	15
2.7.3. Planiranje .....	15
2.7.4. Implementacija i kontrola .....	15

<b>3. RUDARENJE PODATAKA.....</b>	<b>17</b>
3.1. Općenito u rudarenju podataka.....	17
3.1.1. Interpretacija .....	18
3.1.2. Predviđanje .....	18
3.2. Modeli i metode rudarenja podataka.....	19
3.3. Rudarenje podataka, klasična statistika i OLAP analiza.....	19
3.4. Primjena rudarenja podataka .....	20
3.4.1. Relacijski marketing.....	20
3.4.2. Otkrivanje prijevara.....	20
3.4.3. Procjena rizika .....	21
3.4.4. Rudarenje teksta.....	21
3.4.5. Prepoznavanje slika.....	21
3.4.6. Rudarenje weba.....	21
3.4.7. Medicinske dijagnoze.....	21
3.5. Prikaz ulaznih podataka .....	21
3.5.1. Kategorički .....	22
3.5.2. Numerički .....	22
3.5.3. Točke .....	22
3.5.4. Nominalni .....	22
3.5.5. Redni .....	22
3.5.6. Diskretni .....	23
3.5.7. Kontinuirani.....	23
3.6. Proces rudarenja podataka.....	23
3.6.1. Definicija ciljeva .....	23
3.6.2. Prikupljanje i integracija podataka.....	24
3.6.3. Preliminarna analiza .....	25
3.6.4. Odabir atributa .....	25
3.6.5. Razvoj modela .....	25
3.6.6. Interpretacija i predviđanje .....	26
3.7. Analiza metodologija .....	28
3.7.1. Nadzirani procesi učenja .....	28
3.7.2. Nenadzirani procesi učenja.....	28
3.7.3. Karakterizacija i diskriminacija .....	29
3.7.4. Klasifikacija.....	29
3.7.5. Regresija .....	29
3.7.6. Analiza vremenskih nizova.....	30
3.7.7. Asocijativna pravila.....	30
3.7.8. Grupiranje .....	30
3.7.9. Opis i vizualizacija .....	30
<b>4. PRIMJENA SUSTAVA POSLOVNE INTELIGENCIJE U PROIZVODNJI.....</b>	<b>32</b>
4.1. Proizvodni procesi i problemi u proizvodnji .....	32
4.2. Uvođenje sustava poslovne inteligencije u proizvodne organizacije .....	34
4.2.1. Sustavi poslovne i proizvodne inteligencije .....	34
4.2.2. Značaj podataka u proizvodnji.....	35
4.2.3. Prednosti sustava poslovne inteligencije u proizvodnji .....	35
4.2.4. Definiranje ciljeva projekta .....	36



---

4.2.5. Implementacija sustava poslovne inteligencije.....	37
4.3. Poslovna inteligencija u proizvodnji.....	38
4.3.1. Proizvodnja poluvodiča .....	38
4.3.2. Proizvodnja cementa .....	38
4.3.3. Proizvodnja kemikalija .....	39
4.3.4. Proizvodnja elektronskih komponenata .....	39
<b>5. RUDARENJE PODATAKA POMOĆU PROGRAMSKOG SUSTAVA RAPIDMINER .....</b>	<b>40</b>
5.1. Opis baze podataka .....	40
5.2. Model za pronalaženje asocijativnih pravila .....	42
5.2.1. Predprocesuiranje baze podataka za izradu modela asocijativnih pravila.....	42
5.2.2. Izrada modela za kreiranje asocijativnih pravila .....	49
5.2.3. Analiza modela i parametara operatora.....	51
5.2.4. Analiza dobivenih asocijativnih pravila .....	56
5.3. Modeliranje poslovnih pravila.....	58
5.3.1. Izrada prediktivnog modela pomoću neuronskih mreža .....	60
5.3.2. Analiza kreiranog modela i dobivenih rezultata .....	63
5.3.3. Modeliranje poslovnih pravila pomoću induktivnog učenja .....	67
5.3.4. Prediktivni model s induktivnim učenjem.....	74
5.4. Usporedba rezultata modela .....	78
5.5. Zapisivanje i spremanje modela .....	79
<b>6. ZAKLJUČAK .....</b>	<b>80</b>
<b>LITERATURA .....</b>	<b>82</b>
<b>PRILOZI.....</b>	<b>83</b>



## POPIS SLIKA

Slika 1. Prednosti sustava poslovne inteligencije [1] .....	4
Slika 2. Struktura sustava poslovne inteligencije [1] .....	8
Slika 3. Razine sustava poslovne inteligencije [1] .....	9
Slika 4. Ciklus analize sustava poslovne inteligencije [1] .....	10
Slika 5. Razvoj sustava poslovne inteligencije [1] .....	14
Slika 6. Dostupne metodologije sustava poslovne inteligencije [1] .....	16
Slika 7. Proces rudarenja podataka [1] .....	24
Slika 8. Sudionici i njihove uloge u procesu rudarenja podataka[1] .....	27
Slika 9. Općeniti tok proizvodnje .....	33
Slika 10. Prikaz nestrukturirane baze podataka .....	42
Slika 11. Sortirana baza podataka .....	43
Slika 12. Podešavanje alatne trake MS Excel-a .....	44
Slika 13. Pristupanje Visual Basic programskom jeziku .....	44
Slika 14. Sučelje Visual Basic dodatka .....	45
Slika 15. Programski kod Macro modula Sort .....	46
Slika 16. Programski kod Macro modula prazn .....	47
Slika 17. Model za pretvorbu numeričkih vrijednosti .....	48
Slika 18. Postavke „Numerical to Binominal“ operatora .....	48
Slika 19. Kreirani model za generiranje asocijativnih pravila .....	49
Slika 20. Podešavanje <i>Read Excel</i> operatora .....	51
Slika 21. Predloženi tip vrijednosti .....	51
Slika 22. Tip vrijednosti prije obrade podataka .....	52
Slika 23. Tip vrijednosti nakon obrade pomoću operatora „Numerical to Binominal“ .....	52
Slika 24. Postavke „FP- Growth“ operatora .....	53
Slika 25. Generirani frekventni podskupovi .....	53
Slika 26. Postavke operatora za kreiranje asocijativnih pravila .....	54
Slika 27. Dobivena asocijativna pravila .....	56
Slika 28. Asocijativna pravila poredana prema povjerenju .....	57
Slika 29. Izlaz operatora za popunjavanje praznih vrijednosti .....	61
Slika 30. Postavljanje oznake na faktoru cijene .....	61
Slika 31. Operator za unakrsnu validaciju modela .....	62
Slika 32. Postavke Neural Net operatora .....	63
Slika 33. Kreirani glavni proces .....	64
Slika 34. Trening i test podproces .....	64
Slika 35. Vektor performansi .....	65
Slika 36. Predviđene vrijednosti uz pomoć modela .....	66
Slika 37. Struktura primijenjene neuronske mreže za treniranje modela za predviđanje faktora cijene .....	66
Slika 38. Težine izlaznih neurona .....	67
Slika 39. Mogućnosti Rule Induction operatora .....	68
Slika 40. Operator za provođenje diskretizacije .....	69
Slika 41. Glavni proces .....	69

---

Slika 42. Trening i test podprocesi.....	69
Slika 43. Vektor performansi modela (1) .....	70
Slika 44. Vektor performansi modela (2) .....	71
Slika 45. Generirana poslovna pravila.....	72
Slika 46. Glavni proces.....	74
Slika 47. Trening i test podprocesi.....	75
Slika 48. Preciznost modela (1) .....	75
Slika 49. Preciznost modela (2) .....	76
Slika 50. Vektor performansi neuronskih mreža.....	78
Slika 51. Vektor performansi modela za induktivno učenje.....	79
Slika 52. Operator za zapisivanje modela.....	79

## POPIS TABLICA

Tablica 1. Razlike između OLAP-a, klasične statistike i rudarenja podataka .....	19
Tablica 2. Opis polja na radnom listu NAL.PODNAL.TEHNO .....	41
Tablica 3. Radni list NAL.PODNAL.TEHNO.DATUM. ....	41
Tablica 4. Radni list PODNAL.TEHNO .....	41
Tablica 5. Radni list PODNAL.TEHNO.TIPRM .....	41
Tablica 6. Asocijativna pravila i vrijednosti značajnosti pravila .....	57

## **POPIS KRATICA**

<b>Naziv</b>	<b>Značenje</b>
BI	Business Intelligence
ERP	Enterprise Resource Planning
ETL	Extract, Transform, Load (ETL alati)
MI	Manufacturing Intelligence
OLAP	On-Line Analytical Processing
FP Growth	Frequent Pattern Growth

## SAŽETAK

U radu su prikazane teorijske osnove metoda poslovne inteligencije s posebnim naglaskom na metode i modele rudarenja podataka. Prikazani su osnovni uvjeti za provedbu analize metodama rudarenja podataka. Testirana je hipoteza kako podaci dostupni u bazi koji su prikupljeni u procesu proizvodnje utječu na faktor cijene proizvoda. Provedenom analizom pokazalo se kako je moguće kreirati modele zadovoljavajuće preciznosti pomoću kojih je menadžment poduzeća u mogućnosti donositi brže i kvalitetnije poslovne odluke. Hipotezom se želi pokazati kako je uz pomoć prediktivnih modela moguće planirati materijalne i financijske potrebe, kao i dati potencijalnim kupcima preporučenu cijenu za izradu traženoga proizvoda. Također su modelirana i poslovna pravila kojima je cilj pomoći u boljoj organizaciji poslovanja poduzeća.

Ključne riječi: poslovna inteligencija, rudarenje podataka, maloserijska proizvodnja

## **SUMMARY**

In this thesis the basic methods of business intelligence are shown. The special focus was on the methods of data mining. The basic conditions for data mining analysis which must be met are shown. The hypothesis was made that the data given in the dataset of a small-series production company has a significant effect on the final price of the product. The analysis was made and the given results had sufficient precision. Thanks to the given results, the company can predict the average cost of various products with certain reliability. Also, the company is able to make faster and better business decisions that are based on scientific methods. The hypothesis had shown that with the help of the prediction models the company will be able to plan enterprise resources, as well as the financial resources. In second part, the business rules are discussed. The created business rules can be used to help better the business organization.

Key words: business intelligence, data mining, small-series production

## 1. UVOD

U ovom diplomskom zadatku želi se analizirati učinkovitost alata poslovne inteligencije u maloserijskom proizvodnom poduzeću. Želi se provjeriti hipoteza kako alati i metode poslovne inteligencije služe kao potpora procesu donošenja odluka te kolike su mogućnosti primjene alata poslovne inteligencije i rudarenja podataka u maloserijskoj proizvodnji. Poseban osvrt u radu je dan na metode i alate rudarenja podataka. U proizvodnom poduzeću je prisutan informacijski sustav uz pomoć kojega poduzeće bilježi i sprema sve podatke o proizvodnji. Problem se javlja prilikom interpretacije i analize podataka. Kroz vrijeme se stvara velika količina podataka čije je tumačenje u cilju donošenja neke poslovne odluke problematično. Rezultat je sporiji odziv poduzeća prema potencijalnom kupcu. Pomoću metoda poslovne inteligencije i rudarenja podataka želi se provjeriti mogu li te metode poduzeću dati više različitih alternativa i kakav rezultat može poduzeće očekivati. U radu je korištena baza podataka proizvodnog poduzeća u kojoj su dostupni podaci prikupljeni u procesu proizvodnje. U bazi je prisutna velika količina podataka koju je metodama poslovne inteligencije i rudarenja podatka potrebno analizirati kako bi se u konačnici postiglo bolje tumačenje dostupnih podataka iz proizvodnje što bi trebalo rezultirati bržim i kvalitetnijim donošenjem poslovnih odluka. Metodama rudarenja podataka cilj je provjeriti može li se modeliranjem poslovnih pravila, asocijativnih pravila te prediktivnih modela dokazati kako metode rudarenja podataka služe kao potpora procesu donošenja bržih i kvalitetnijih poslovnih odluka. Problem je postavljen pomoću baze podataka koju je bilo potrebno analizirati, predprocesuirati te primijeniti modele rudarenja podataka. Cilj je primjenom metoda rudarenja podataka poduzeću omogućiti bolju reakciju na zahtjeve kupaca, čime se povećava konkurentnost poduzeća. Predprocesuiranje je postupak koji se provodi u MS Excel i RapidMiner programskim sustavima, a cilj je dovođenje baze podataka u oblik koji kasnije omogućava analizu baze podataka. Primjenom metoda rudarenja podataka cilj je dobivanje jasnih poslovnih pravila koja omogućavaju bržu reakciju na zahtjeve tržišta i asocijativnih pravila uz pomoć kojih poduzeće može provjeriti koji proizvodi ovise jedni o drugima.



## 2. UVOD U POSLOVNU INTELIGENCIJU

Velike promjene koje su se dogodile u načinu proizvodnje i gospodarskim odnosima dovele su do rasta važnosti razmjene nematerijalne robe, koja se većim dijelom sastoji od prijenosa informacija. Na to utječu dva važna faktora:

- Prvi je globalizacija, odnosno stalno rastuća međuovisnost gospodarstava pojedinih zemalja; stvara se jedinstveno globalno tržište sa visokim stupnjem integracije
- Drugi faktor su nove informacijske tehnologije, koje predvodi masovno širenje Interneta i bežičnih uređaja. Omogućava se prijenos velike količine podataka te raznolikost u primjeni sredstava komunikacije [1, 2, 3]

Jednostavan pristup informacijama i znanju nudi razne prednosti sudionicima društveno ekonomskog okruženja. Pojedinci mogu puno brže doći do novih saznanja i informacija o događanjima, ili mogu obavljati trgovačke i bankarske transakcije putem Interneta. Poduzeća koja razvijaju nove proizvode i usluge mogu uspješnije zadovoljiti potrebe korisnika, ili steći konkurentsku prednost učinkovitijim korištenjem stečenog znanja. Pojava jeftinih tehnologija za masivnu pohranu podataka i rasprostranjenost Interneta omogućila je raznim organizacijama da tijekom godina prikupe i pohrane veliku količinu podataka. Poduzeća koja su sposobna transformirati te podatke u znanja mogu ih iskoristiti za donošenje učinkovitijih odluka i tako steći konkurentsku prednost.

Poslovna inteligencija (eng. *business intelligence*) se definira kao skup matematičkih modela i metodologija analiza koje sustavno koriste raspoložive podatke za dobivanje informacija i znanja korisnih za složene procese donošenja odluka. [1, 2, 3]

### 2.1. Značaj učinkovitih i pravovremenih odluka

U složenim organizacijama odluke se donose na dnevnoj bazi. Te odluke mogu biti više ili manje važne, mogu imati dugotrajan ili kratkoročan učinak, i mogu uključivati ljude na raznim hijerarhijskim razinama. Sposobnost za donošenje odluka, bilo pojedinca ili zajednice, je jedan od primarnih faktora koji utječu na učinkovitost i konkurentnost organizacije. [1, 2, 3, 4, 5]

Većina odluka se donosi koristeći jednostavne i intuitivne metode, koje uzimaju u obzir specifične elemente kao što su iskustvo, znanje i dostupne informacije. Takav pristup dovodi do stagnacije načina donošenja odluka, koji je neprikladan za nestabilne uvijete nametnute učestalim i brzim gospodarskim promjenama. U današnjim organizacijama, procesi odlučivanja su često presloženi i dinamični da bi se radilo na intuitivan način. Umjesto toga zahtijevaju stroži pristup temeljen na analitičkim metodama i matematičkim modelima.

Glavna uloga sustava poslovne inteligencije je pružanje alata i metoda koje omogućuju donošenje učinkovitih i pravodobnih odluka. [1, 2, 3, 4, 5]

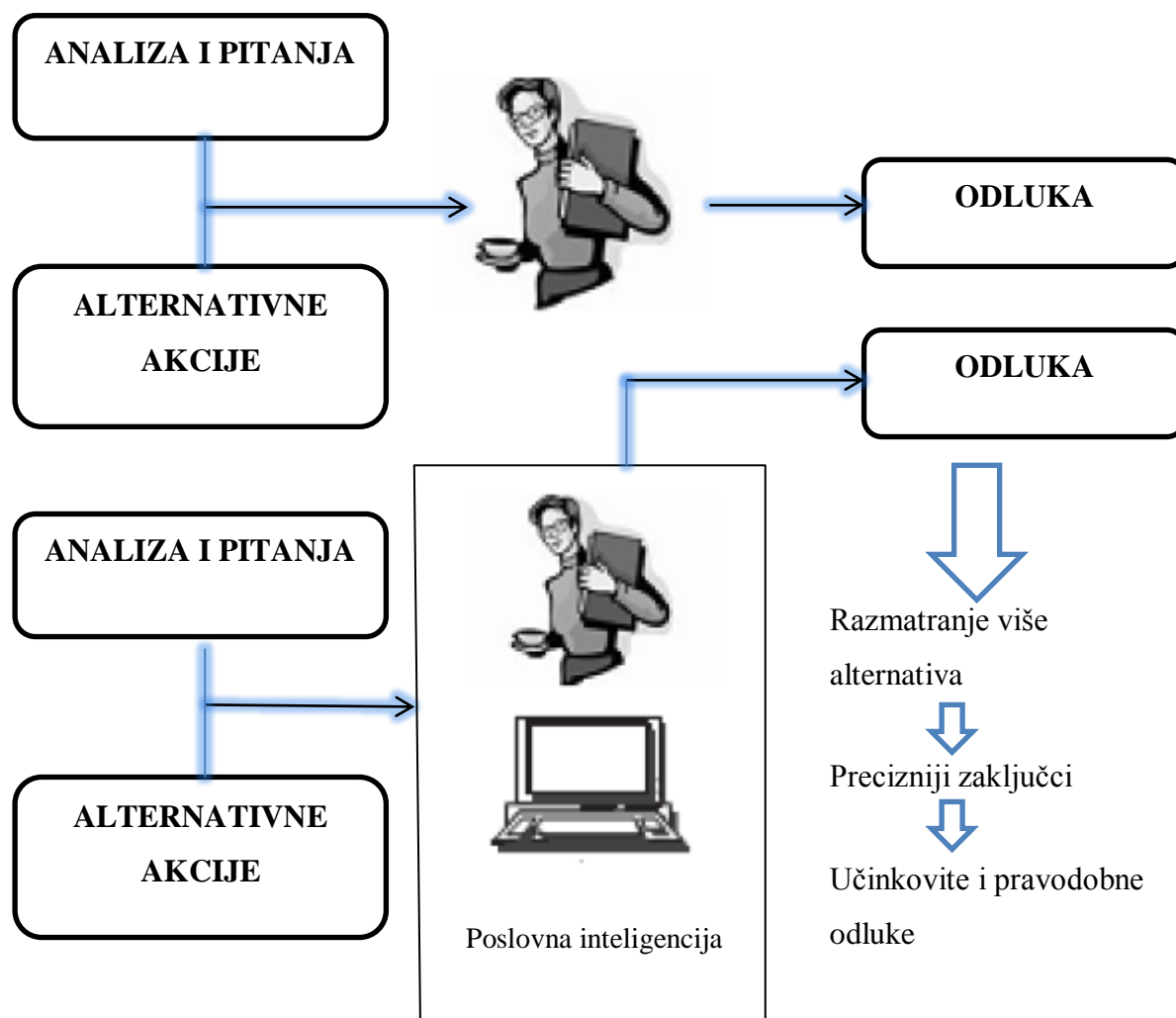
### **2.1.1. Učinkovite odluke**

Primjena strogih analitičkih metoda omogućuje donositeljima odluka da se oslanjaju na informacije i znanja koja su pouzdanija. Rezultat toga je donošenje boljih odluka i planova djelovanja koji omogućuju učinkovitije postizanje ciljeva. Međutim, analitičke metode zahtijevaju eksplicitan opis kriterija za procjenu alternativa i mehanizama koji reguliraju problem. Osim toga, nužan je temeljni pregled i razumijevanje temeljne logike procesa donošenja odluka.

### **2.1.2. Pravovremene odluke**

Poduzeća posluju u gospodarskom okruženju koje karakterizira rast razine konkurencije i visoka dinamičnost. Zato je sposobnost da se brzo reagira na akcije konkurencije i na nove uvijete tržišta kritični faktor za uspjeh ili čak opstanak tvrtke.

Slika 1 pokazuje glavne dobiti koje organizacija može izvući uvođenjem sustava poslovne inteligencije. Kada se donosioci odluka suočavaju s problemima postavljaju si niz pitanja i razvijaju odgovarajuće analize. Nekoliko opcija se ispituje i uspoređuje, te se s obzirom na uvjete koji su uzeti u obzir, odabire najbolja. Ako se donosioci odluka mogu osloniti na sustav poslovne inteligencije, mogu se očekivati velika poboljšanja cjelokupne kvalitete procesa donošenja odluka.



**Slika 1. Prednosti sustava poslovne inteligencije [1]**

Pomocu matematičkih modela i algoritama, moguće je analizirati veći broj alternativnih akcija, dolazi se do boljih zaključaka, te učinkovitijih i pravodobnih odluka. Stoga je moguće zaključiti da je glavna prednost koja proizlazi iz usvajanja sustava poslovne inteligencije povećanje učinkovitosti procesa donošenja odluka.

## 2.2. Podaci, informacije i znanje

U informacijskim sustavima, javnih i privatnih organizacija, nakupljaju se velika količina podataka. Ti podaci potječu dijelom od internih transakcija administrativne, logističke i komercijalne naravi, te dijelom od vanjskih izvora. Međutim, čak i ako su prikupljeni i pohranjeni na sustavan način, ne mogu se izravno koristiti za donošenje odluka. Podaci trebaju biti obrađeni odgovarajućim alatima za ekstrakciju i analitičkim metodama,

sposobnim transformirati ih u informacije i znanja, koja bi se mogla koristiti za donošenje odluka.

Slijedi objašnjenje razlike između podataka, informacija i znanja.

### **2.2.1. Podaci**

Podaci uglavnom predstavljaju kodifikaciju primarnog entiteta, kao i transakcije koje uključuju dva ili više primarnih entiteta. [1, 2, 3, 4, 5] Primjerice, za trgovca primarni entiteti mogu biti kupci, prodajna mjesta i roba koja se prodaje, dok račun predstavlja komercijalnu transakciju.

### **2.2.2. Informacije**

Informacije su rezultat ekstrakcije i obrade podataka, i imaju značenje onome kome su potrebne. [1, 2, 3, 4, 5] Primjerice, voditelju prodaje maloprodajnog poduzeća, udio računa s iznosom većim od 100 novčanih jedinica u jednome tjednu, predstavlja značajnu informaciju koja može biti izvađena iz sirovih podataka.

### **2.2.3. Znanje**

Informacija se pretvara u znanje kada se koristi za donošenje odluka i planiranje odgovarajućih akcija. [1, 2, 3, 4, 5] Znanjem se smatra skup informacija iz nekog područja, potpomognutih iskustvom i kompetencijom donositelja odluka u rješavanju složenih problema. Analiza prodaje maloprodajnog poduzeća, može otkriti da je grupa kupaca koja živi na području, gdje je konkurent otvorio prodajno mjesto, smanjila potražnju. S vremenom će znanje prikupljeno na ovaj način dovesti do akcije koja će biti usmjerena na rješavanje problema, primjerice uvođenjem usluge besplatne dostave za kupce na tom području. Znanje može biti prikupljeno iz podataka na pasivan način, analitičkim kriterijem predloženim od strane donositelja odluka, ili aktivnom primjenom matematičkih modela, u obliku induktivnog učenja ili optimizacije.

Postoje javna i privatna poduzeća koja su u proteklih nekoliko godina razvila formalne i sustavne mehanizme prikupljanja, pohrane i podjele znanja. Mehanizme prikupljanja sada smatraju neprocjenjivom nematerijalnom imovinom. Aktivnosti pružanja podrške u širenju znanja, kroz organizaciju, integracijom procesa donošenja odluka i usvajanjem informacijskih tehnologija obično se nazivaju upravljanje znanjem.

Očito je da poslovna inteligencija i upravljanje znanjem dijele neke sličnosti u svojim ciljevima. Glavni cilj objiju disciplina je razviti okruženje koje podupire donošenje odluka u procesu donošenja odluka i aktivnostima rješavanja složenih problema. Da bi se uočila razlika između ove dvije discipline, potrebno je primijetiti da se metode upravljanja znanjem primarno orijentiraju na obradu informacija koje su obično nestrukturirane, ponekad implicitne, i uglavnom se nalaze u dokumentima, razgovorima i iskustvima. Poslovna inteligencija se temelji na strukturiranim informacijama, najčešće kvantitativne prirode i obično organizirane u baze podataka. Međutim, ta razlika je pomalo nejasna. Primjerice, analizom e-mail adresa i Internet stranica metodama rudarenja podataka sustavi poslovne inteligencije počinju se baviti nestrukturiranim podacima.

### **2.3. Uloga matematičkih modela**

Sustavi poslovne inteligencije pružaju donositeljima odluka informacije i znanja izvađena iz podataka, primjenom matematičkih modela i algoritama. U nekim se slučajevima ova aktivnost može svesti na izračun postotka i grafički prikaz jednostavnih histograma, dok složenije analize zahtijevaju razvoj naprednih modela optimizacije i učenja. [1, 2, 3, 4, 5]

Klasične znanstvene discipline, poput fizike, uvijek su posezale za matematičkim modelima da bi opisale realne sustave. Druge discipline, kao što su operacijska istraživanja, su koristile znanstvene metode i matematičke modele za proučavanje umjetnih sustava, primjerice privatnih i javnih organizacija.

Generalno gledano, usvajanje sustava poslovne inteligencije teži promicanju znanstvenog i racionalnog pristupa upravljanju poduzeća i složenih organizacija. Racionalni pristup sustava poslovne inteligencije može se podijeliti po glavnim koracima. Najprije se definiraju objekti analize i pokazatelji performansi koji će se koristiti za evaluaciju alternativnih mogućnosti. Zatim se razvija matematički model koji se temelji na vezama između varijabli, parametrima i mjerama evaluacije zadanog sustava. Naposljetku se provodi what-if analiza, kojom se vrši evaluacija performansi koje ovise o varijaciji varijabla i promjenama parametara sustava.

Razvoj apstraktnog modela tjera donosioca odluka da se usredotoče na glavne značajke analiziranih sustava, što dovodi do boljeg razumijevanja danog sustava. Znanje o sustavu, koje je prikupljeno prilikom izrade matematičkog modela, moguće je jednostavno prenijeti svim pojedincima unutar organizacije. Na taj je način omogućeno bolje očuvanje

znanja u odnosu na empirijski proces donošenja odluka. Matematički modeli koji su razvijeni za specifičan slučaj, vrlo su općeniti i fleksibilni da se u većini slučajeva mogu koristiti u rješavanju sličnih problema. [1, 2, 3, 4, 5]

## **2.4. Arhitektura sustava poslovne inteligencije**

Arhitektura sustava poslovne inteligencije opisana je na slici 2., i uključuje tri glavne komponente.

### **2.4.1. Izvori podataka (eng. *Data sources*)**

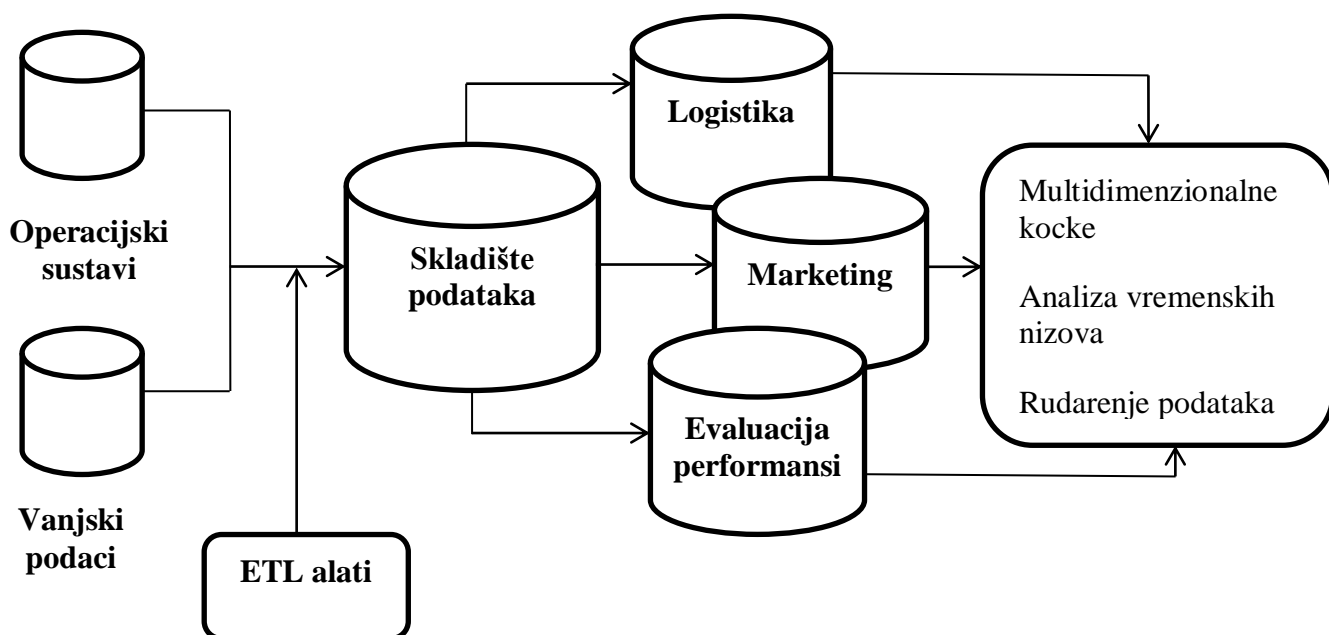
Najprije se sakupljaju i integriraju podaci, različite vrste i podrijetla, pohranjeni u raznim izvorima. Izvori se uglavnom sastoje od podataka koji pripadaju operacijskim sustavima, ali mogu sadržavati nestrukturirane dokumente kao što je elektronička pošta i podaci dobiveni od vanjskih dobavljača. Potrebni su veliki naponi da se ujedine i integriraju različiti izvori podataka.

### **2.4.2. Skladišta podataka (eng. *Data warehouses and Data marts*)**

Koriste se alati za vađenje, transformaciju i učitavanje podataka (eng. *extract, transform and load, ETL*), da bi se podaci iz različitih izvora sortirali u baze podataka (skladišta podataka), odnosno da bi se omogućile analize pomoću alata poslovne inteligencije.

### **2.4.3. Metodologije poslovne inteligencije**

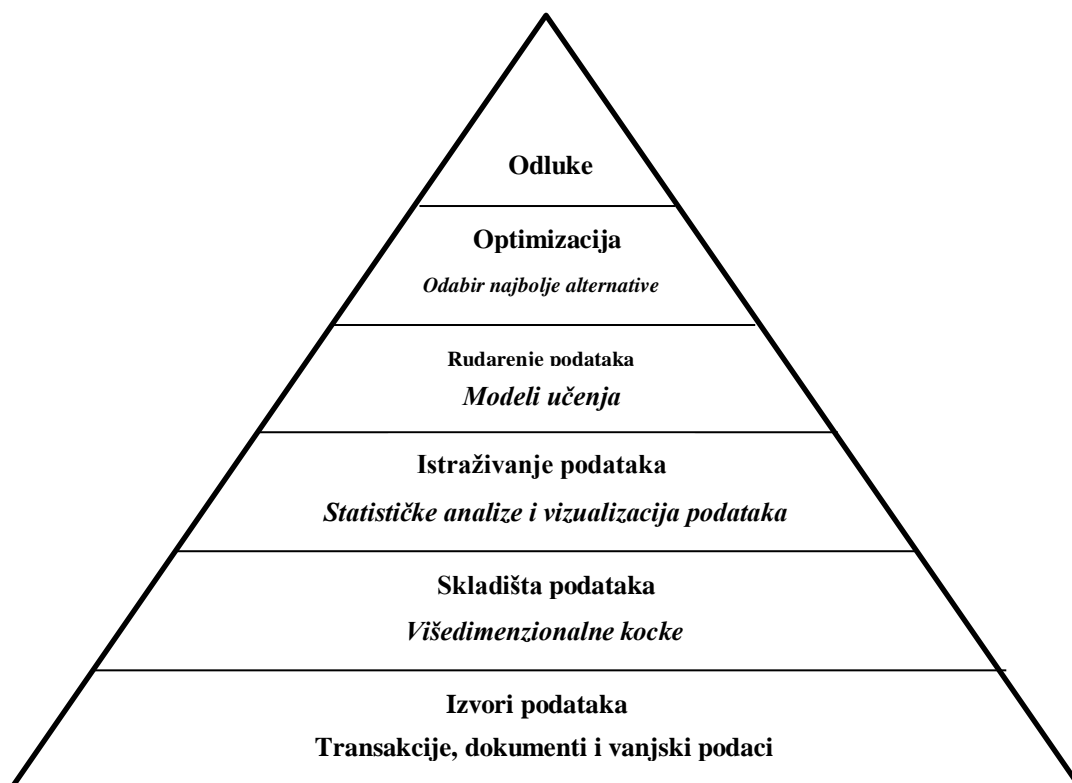
Nakon što su podaci pohranjeni u spremišta podataka, dostupni su matematičkim modelima i analizama namijenjenim za podršku donosiocima odluka. U sustave poslovne inteligencije, ugrađene su neke aplikacije za podršku donošenja odluka. Neke od njih su: analiza multidimenzionalne kočke, analiza vremenskih nizova, induktivni modeli učenja za rudarenje podataka (eng. *data mining*), modeli za optimizaciju.



Slika 2. Struktura sustava poslovne inteligencije [1]

Piramida na slici 3 prikazuje razine sustava poslovne inteligencije. Na slici 2 prikazane su prve dvije razine. Slijedi opis ostalih razina.





Slika 3. Razine sustava poslovne inteligencije [1]

#### 2.4.4. Istraživanje podataka

Na trećoj razini piramide, nalaze se alati za provedbu pasivnih analiza sustava poslovne inteligencije, koji se sastoje od upita (eng. *query*), sustava izvješćivanja (eng. *reporting systems*) i statističkih metoda. Nazivaju se pasivnima zato jer je donosioci odluka moraju odrediti hipoteze ili kriterije izvlačenja podataka, i onda koristiti alate za analizu. Primjerice, voditelj prodaje primijeti da su prihodi određene grupe kupaca na određenom području smanjeni. Voditelj bi možda poželio, uporabom alata za ekstrakciju i vizualizaciju, te statističkim testom provjeriti da li su njegovi zaključci popraćeni adekvatnim podacima.

#### 2.4.5. Rudarenje podataka

Na četvrtoj razini nalaze se aktivne metodologije sustava poslovne inteligencije, kojima je svrha vađenje informacija i znanja iz podataka. Te metodologije uključuju matematičke modele raspoznavanja, strojno učenje i tehnike rudarenja podataka. Za razliku od alata navedenih u nižim razinama piramide, ovi su modeli aktivni, te donosioci odluka ne moraju formulirati hipotezu, koja će se kasnije provjeravati. Njihova svrha je proširiti znanje donosiocima odluka.

### 2.4.6. Optimizacija

Na sljedećoj razini piramide nalazi se model optimizacije koji omogućuje odabir najboljeg rješenja od mogućih alternativa. Uglavnom je skup alternativa prilično opsežan, a ponekad i beskonačan.

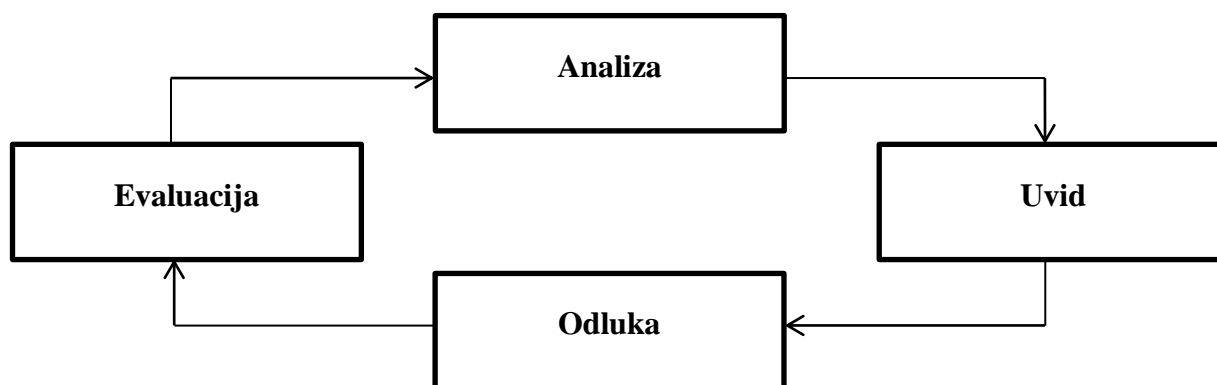
### 2.4.7. Odluka

Na vrhu piramide nalazi se odluka, koja predstavlja zaključak procesa donošenja odluka. Međutim, iako su metodologije poslovne inteligencije dostupne i uspješno usvojene, odluku moraju donijeti donosioci odluka. Donosioci odluka mogu koristiti neformalne i nestrukturirane informacije koje su im dostupne, te na temelju njih modificirati ili promijeniti odluku preporučenu od strane matematičkih modela.

Krećući se od dna prema vrhu piramide, sustavi poslovne inteligencije nude znatno snažnije aktivne alate. Moguća je promjena stručnosti i nadležnosti zaposlenika. Na dnu piramide, potrebni su stručnjaci za informacijske sustave, koji se obično nazivaju administratorima baza podataka. Za matematičke i statističke modele, koji se nalaze u srednjim razinama piramide, odgovorni su analitičari i stručnjaci. Za aktivnosti na vrhu piramide zaduženi su donosioci odluka.

## 2.5. Ciklus analize sustava poslovne inteligencije

Svaka analiza pomoću sustava poslovne inteligencije slijedi svoj tok ovisno o području primjene, subjektivnosti donosioca odluka, te dostupnih analitičkih metodologija. Međutim, moguće je definirati idealan ciklus analize sustava poslovne inteligencije, kao što je prikazano na slici 4.



Slika 4. Ciklus analize sustava poslovne inteligencije [1]

### **2.5.1. Analiza**

Tijekom analize prepoznaje se zadani problem i identificiraju se kritični faktori koji se doimaju najvažnijim. Primjena metodologija poslovne inteligencije omogućuje brz razvoj različitih modela istraživanja, dok se ne pronađe zadovoljavajući model.

### **2.5.2. Uvid**

Druga faza donosiocima odluka pruža mogućnost boljeg razumijevanja danog problema. Primjerice, ako provedena analiza pokazuje da velik broj klijenata zatvara policu osiguranja nakon isteka godišnjeg ugovora, onda se u ovoj fazi identificiraju sličnosti koje ti klijenti dijele. U ovoj su fazi, informacije prikupljene u fazi analize, pretvorene u znanje. Znanje može biti stečeno intuicijom donositelja odluka, koja se temelji na njegovom iskustvu ili nestrukturiranim informacijama koje su im dostupne. S druge strane, induktivni modeli učenja pokazali su se vrlo korisnim u ovoj fazi, naročito ako su primijenjeni na strukturirane podatke.

### **2.5.3. Odluka**

Tijekom ove faze, znanje stečeno u fazi uvida pretvara se u odluku, a zatim u akciju. Pomoću metodologija poslovne inteligencije moguće je brže izvoditi fazu analize i uvida, kako bi se postigle bolje i pravovremene odluke. Time je smanjeno vrijeme ciklusa analize, odluke, akcije i revizije, te je i donošenje odluka kvalitetnije.

### **2.5.4. Evaluacija**

Posljednja faza ciklusa poslovne inteligencije uključuje mjerenja i evaluaciju. Smišljaju se opsežna mjerenja, koja ne uključuju samo financije, već se i ključni pokazatelji uspješnosti različitih odjela organizacije uzimaju u obzir.

## **2.6. Ključni faktori razvoja sustava poslovne inteligencije**

Faktori koji utječu na uspjeh projekta sustava poslovne inteligencije dijele se na tehnološke, analitičke i ljudske resurse. Neki su faktori važniji od drugih, za uspjeh projekta. [1, 2, 3]

### **2.6.1. Tehnologija**

Hardverske i softverske tehnologije bitne su faktori za razvoj sustava poslovne inteligencije poduzeća i složenih organizacija. Tijekom posljednja dva desetljeća, snaga mikroprocesora povećava se 100% svakih 18 mjeseci, a i cijena im je pada. Taj trend omogućuje primjenu naprednih algoritama koji koriste metode induktivnog učenja i modele optimizacije, te se vodi računa o njihovom vremenu izvođenja. Također je moguće koristiti tehnike vizualizacije, s prikazima u realnom vremenu. Sljedeći ključni faktor je eksponencijalno povećanje kapaciteta uređaja za pohranu podataka, čija cijena također pada, što omogućuje organizacijama pohranu velikog broja podataka u sustave poslovne inteligencije. Tu je i mrežno povezivanje, odnosno ekstranet (eng. *extranet*) i intranet (eng. *intranet*), koji imaju jednu od glavnih uloga u širenju informacija i znanja unutar organizacije. Na posljetku, jednostavnost integracije hardvera i softvera različitih proizvođača, bitan je faktor koji utječe na alate za analizu podataka.

### **2.6.2. Analiza**

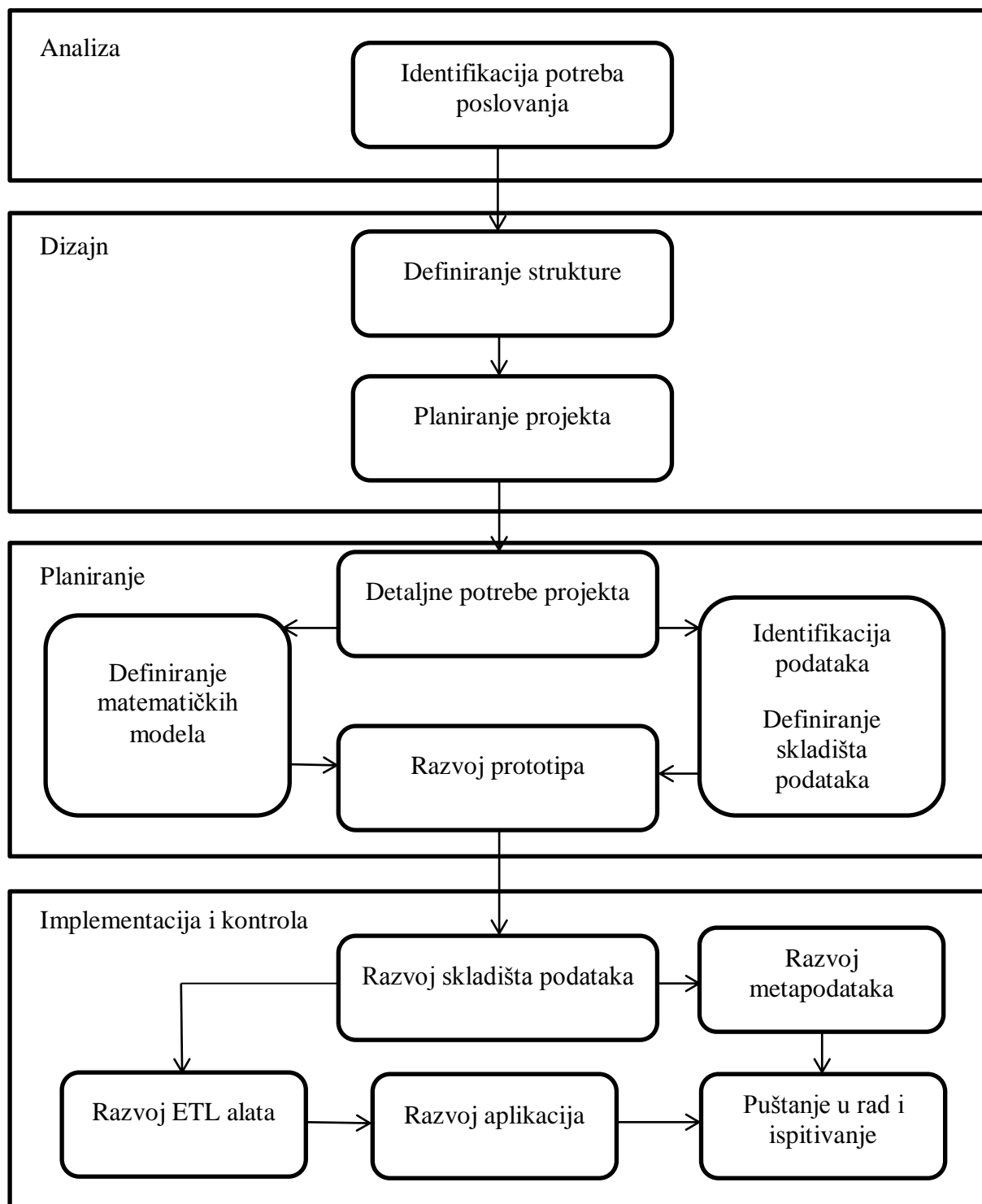
Matematički modeli i analitičke tehnologije imaju ključnu ulogu u vađenju informacija i znanja iz dostupnih podataka. Sama vizualizacija podataka, pasivan je oblik podrške, ali ima važnu ulogu u olakšavanju procesa donošenja odluka. Zato se primjenjuje više naprednih modela induktivnog učenja i optimizacije, s ciljem postizanja aktivnog oblika potpore procesu donošenja odluke.

### **2.6.3. Ljudski potencijali**

Vrijednost ljudskih potencijala čine ljudi koji djeluju unutar organizacije, bilo kao pojedinci ili timovi. Ukupna znanja koja posjeduju i dijele pojedinci čine organizacijsku kulturu. Sposobnost sakupljanja informacija i pretvaranja u akcije jedan je od glavnih faktora svake organizacije, i ima velik utjecaj na kvalitetu procesa odlučivanja. Ako je poduzeće uvelo napredni sustav poslovne inteligencije, još uvijek su potrebni radnici koji će provoditi analize, interpretirati rezultate, pronalaziti kreativna rješenja i smišljati učinkovite planove djelovanja. Organizacije, čiji su zaposlenici spremni prihvatiti promjene u načinu donošenja odluka, steći će prednost nad konkurencijom.

## **2.7. Razvoj sustava poslovne inteligencije**

Razvoj sustava poslovne inteligencije vrlo je sličan razvoju projekta. Sličan je jer je potrebno definirati konačan cilj sustava, vrijeme i troškove razvoja sustava, te resurse koji su potrebni za izvođenje planiranih aktivnosti. Na slici 2.5 prikazan je tipičan ciklus razvoja sustava poslovne inteligencije. Razvoj sustava poslovne inteligencije može se razlikovati od prikazanog na slici. Primjerice, ako već postoje skladišta podataka sa osnovnom informacijskom strukturom, tada odgovarajuće faze sa slike 5 neće biti potrebne.



Slika 5. Razvoj sustava poslovne inteligencije [1]

### 2.7.1. Analiza

Utvrđuju se potrebe organizacije koje su bitne za sustav poslovne inteligencije. Ta se faza obično provodi kroz niz intervjuja sa zaposlenicima u organizaciji. Bitno je jasno opisati

glavne ciljeve i prioritete sustava, kao i troškove i prednosti uvođenja sustava poslovne inteligencije.

### **2.7.2. Dizajn**

Ova faza se sastoji od dvije manje faze kojima je cilj odrediti privremeni plan cjelokupne arhitekture, uzimajući u obzir razvoj u bliskoj budućnosti. Najprije se radi procjena postojećih informacijskih struktura. Da bi se prikupile sve potrebne informacije, proučava proces donošenja odluka koji će sustav poslovne inteligencije podržavati. Zatim se pomoću klasičnih metodologijama projektnog menadžmenta iznosi plan, utvrđene faze razvoja, prioriteti, očekivano vrijeme razvoja i troškovi, te potrebni resursi.

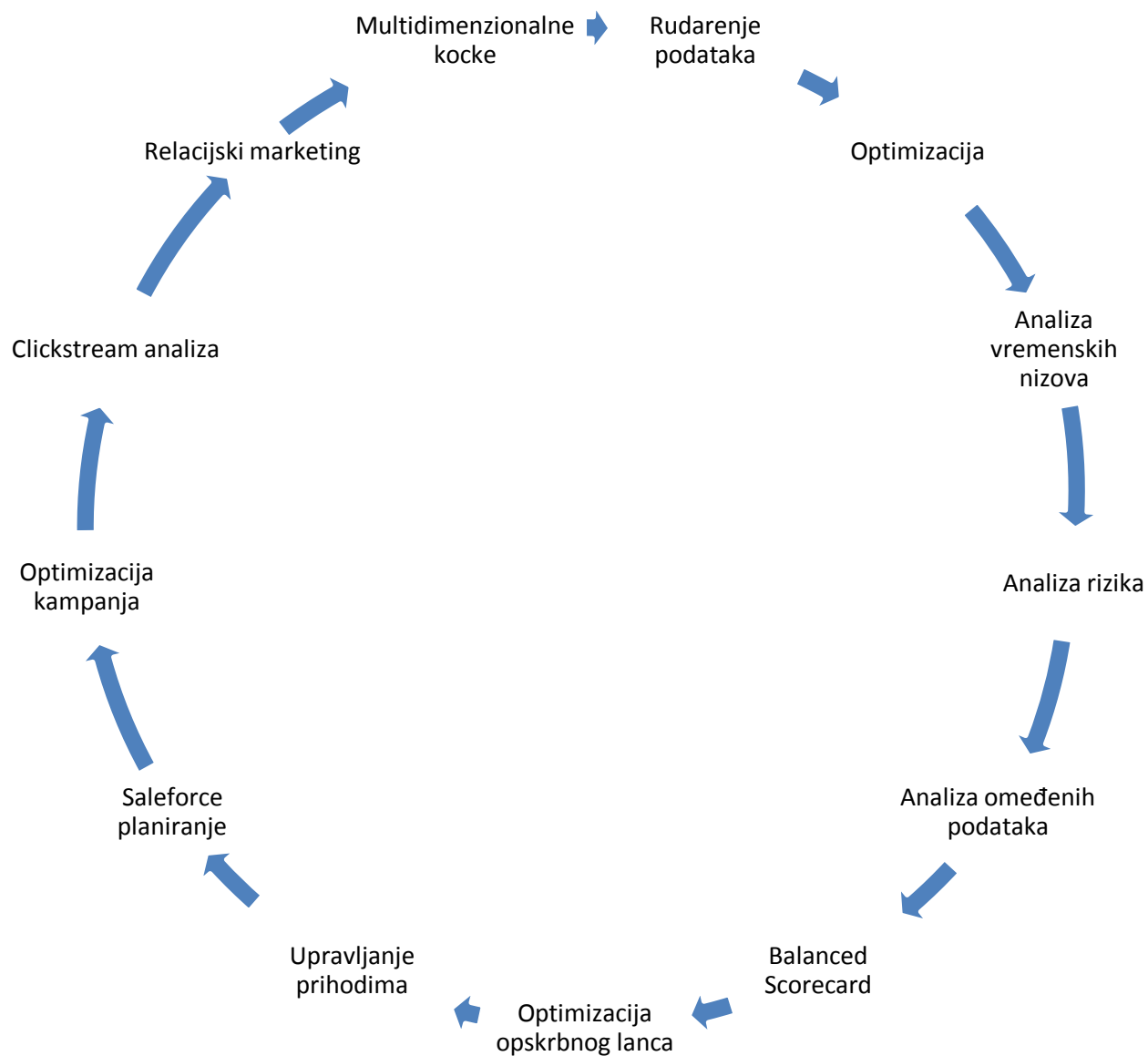
### **2.7.3. Planiranje**

Faza planiranja uključuje podfaze, u kojima su detaljno definirane i opisane funkcije sustava poslovne inteligencije. Procjenjuju postojeći podaci i podaci iz vanjskih izvora. To omogućuje oblikovanje strukture sustava poslovne inteligencije, koja se sastoji centralnog skladišta podataka (eng. *central data warehouse*) i ako je potrebno satelitskih skladišta podataka (eng. *satellite data mart*). Prepoznavanjem raspoloživih podataka, usvajaju se matematički modeli koje je potrebno definirati. Na taj način se osigurava dostupnost podataka za sve algoritme koji će biti korišteni za rješavanje problema. Zadnji korak ove faze je izrada prototipa, niske cijene i ograničenih mogućnosti, s ciljem otkrivanja različitosti između stvarnih potreba i onih definiranih u projektu.

### **2.7.4. Implementacija i kontrola**

Posljednja faza se također sastoji od podfaza. Razvijaju se skladišta podataka. Ona predstavljaju informacijsku strukturu koja će sustav poslovne inteligencije snabdijevati s informacijama. Objašnjenja značenja podataka koji se nalaze u skladištima podataka, zapisuju se u dokument koji se naziva arhiva metapodataka (eng. *metadata*). Utvrđuju se ETL procedure koje će vaditi i transformirati podatke iz primarnih izvora, te ih unositi u skladišta podataka. Razvijaju se ključne aplikacije sustava poslovne inteligencije koje omogućuju provođenje planiranih analiza. U zadnjoj podfazi slijedi ispitivanje i korištenje sustava. Slika 6 prikazuje glavne metodologije koje mogu biti korištene u sustavima poslovne inteligencije. Neke od njih mogu biti korištene u raznim područjima, dok se ostale koriste samo za određene zadatke.





Slika 6. Dostupne metodologije sustava poslovne inteligencije [1]

### 3. Rudarenje podataka

Cilj aktivnosti uključenih u analize velikih baza podataka je vađenje korisnih informacija za podršku procesa donošenja odluka. Te se aktivnosti nazivaju: rudarenje podataka, raspoznavanje uzoraka, i strojno učenje. [1, 2, 3, 4, 5] Rudarenje podataka je proces istraživanja i analize skupa podataka, koji je obično velik, s ciljem pronalaženja odgovarajućih obrazaca koji se ponavljaju, kako bi se dobilo potrebno znanje. Uloga rudarenja podataka sve je važnija za teoretska istraživanja, kao i za konkretnu primjenu.

U ovom poglavlju je opisano i karakterizirano rudarenje podataka s obzirom na svrhu istraživanja i metodologije analiza. Također su objašnjena bitna svojstva ulaznih podataka, i opisan je proces rudarenja podataka.

#### 3.1. Općenito u rudarenju podataka

Aktivnosti rudarenja podataka čine iterativni procesi usmjereni na analizu velikih baza podataka, kako bi se izvadile informacije i znanja koja mogu biti korisna u donošenju odluka. Proces analize je iterativan i ima različite faze koje mogu slati povratne informacije i izvoditi naknadne revizije. Za oblikovanje takvih procesa obično je potrebna suradnja stručnjaka na području u kojem se primjenjuje i analitičara podataka koji koriste matematičke modele za induktivno učenje. Praksa pokazuje da izvođenje rudarenja podataka zahtijeva česte intervencije analitičara tijekom različitih faza, te stoga se ne može lako automatizirati. Nužno je da je izvađeno znanje točno, odnosno da je utemeljeno na podacima, i ne smije dovesti do pogrešnih zaključaka.

Pojam rudarenje podataka odnosi se na sve procese prikupljanja i analiziranja podataka, razvoj modela induktivnog učenja, usvajanje odluka i akcije koje su rezultat stečenog znanja. Pojam teorija matematičkog učenja (eng. *mathematical learning theory*) odnosi se na niz matematičkih modela i metoda koje se mogu naći u jezgri svake analize rudarenja podataka koja se koristi za stvaranje novih znanja.

Proces rudarenja podataka temelji se na metodama induktivnog učenja, čija je glavna svrha izvođenje općih pravila počevši od dostupnih primjera koji se sastoje od zabilježenih promatranja iz prošlosti sačuvanih u jednoj ili više baza podataka. Drugim riječima, svrha

analiza rudarenja podataka je izvući neke zaključke iz opažanja iz prošlosti i generalizirati ih na što je moguće precizniji način. Modeli i obrasci koji se utvrde na taj način mogu biti različitog oblika, primjerice, linearne jednadžbe, skup if-then-else pravila, klasteri i grafovi.

Nadalje, rudarenje podataka ovisi o postupcima prikupljanja zapažanja i unosom istih u bazu podataka, iako primarna svrha pohrane tih podataka nisu analize rudarenja podataka. Primjerice, telefonske kompanije zapise o pozivima pohranjuju za administrativnu upotrebu, dok se kasnije oni mogu koristiti za izvođenje analiza rudarenja podataka. Procedura prikupljanja podataka ne ovisi, niti je u skladu s ciljevima rudarenja podataka, tako da se to prikupljanje razlikuje od prikupljanja podataka koje se provodi za uzimanje uzoraka, po unaprijed određenoj shemi.

Aktivnosti rudarenja podataka mogu se podijeliti u dvije skupine, u klasu s glavnim ciljem analize: interpretacija i predviđanje.

### **3.1.1. Interpretacija**

Cilj interpretacije je identificirati obrasce podataka koji se ponavljaju i zatim ih izraziti pravilima i kriterijima koje će stručnjaci na području primjene lako razumjeti. Pravila moraju biti originalna i netrivialna kako bi se povećala razina znanja i razumijevanja promatranog sustava. Primjerice, za maloprodajnu industriju, moglo bi biti korisno klasificirati kupce, koji posjeduju potrošačke kartice (eng. *loyalty card*), ovisno o profilu kupnje. Na taj se način stvara segment koji bi mogao biti koristan za pronalaženju marketinške niše i usmjeravanje budućih marketinških kampanja.

### **3.1.2. Predviđanje**

Svrha predviđanja je predvidjeti koju će vrijednost slučajna varijabla poprimiti u budućnosti ili procijeniti vjerojatnost budućih događaja. Primjerice, telefonske kompanije mogu razviti analizu rudarenja podataka za procjenu vjerojatnosti prelaska klijenata kod konkurencije. Većina tehnika rudarenja podataka vrši predviđanja na temelju vrijednosti varijabli entiteta u bazi podataka. Primjerice, model rudarenja podataka može ukazivati na to da vjerojatnost prelaska klijenata kod konkurencije ovisi o značajkama kao što su dob, trajanje ugovora i udio poziva na mreže konkurencije.

### 3.2. Modeli i metode rudarenja podataka

Postoji nekoliko metoda učenja za izvođenje različitih zadataka rudarenja podataka. Velik broj tehnika potječe s područja računalne znanosti, kao što je klasifikacija stabala ili asocijativna pravila, i nazivaju se strojno učenje (eng. *machine learning*) ili otkrivanje znanja (eng. *knowledge discovery*). Unutar ove vrste tehnika u većini slučajeva prevladava empirijski pristup. Ostale metode spadaju u multivarijantnu statistiku, kao što je regresija ili Bayesianovi klasifikatori, i često su parametarske i teoretski više utemeljene. U zadnje vrijeme koriste se matematičke metode učenja, kao što je statistička teorija učenja (eng. *statistical learning theory*), koje se temelje na čvrstim teoretskim osnovama i uključuju više disciplina, među kojima se nalaze teorija vjerojatnosti, optimizacije i statistike.

Bez obzira na specifične metode učenja koje se žele usvojiti, postoje koraci za razvoj modela rudarenja podataka:

- odabir klase modela koji će se koristiti za učenje iz prošlosti i specifičnog oblika predstavljanja obrazaca
- definiranje mjera za evaluaciju učinkovitosti i točnosti odabranog modela
- oblikovanje računalnog algoritma, za generiranje modela, optimizirajući mjere evaluacije [1, 2, 3]

### 3.3. Rudarenje podataka, klasična statistika i OLAP analiza

Projekti rudarenja podataka razlikuju se po klasičnoj statistici i OLAP analizama. Razlike su prikazane u tablici 1.

**Tablica 1. Razlike između OLAP-a, klasične statistike i rudarenja podataka**

OLAP	Klasična statistika	Rudarenje podataka
Ekstrakcija detalja i agregiranje podataka	Provjera hipoteze koju je kreirao analitičar	Identifikacija obrazaca podataka
Informacija	Provjera	Znanje
Raspodjela prihoda potraživača zajma za kuću	Analiza varijance prihoda potraživača	Karakterizacija kuće i predviđanje budućih potraživača

Glavna razlika je u aktivnoj orijentaciji koju nude modeli induktivnog učenja, za razliku od statističkih tehnika i OLAP koje su pasivnije. Statističke analize donositelji odlika formuliraju hipotetski, te one trebaju biti potvrđene na temelju uzoraka. Slično tome, na temelju intuicije donositelja odluka OLAP analize temelje ekstrakciju, izvještavanje i kriterije vizualizacije. Statističke metoda i alati za navigaciju kroz kocke podataka koriste elemente za potvrdu ili odbacivanje formuliranih hipoteza, tijekom analize je top-down. Suprotno tome, modeli učenja svojom aktivnom ulogom generiraju predviđanja i interpretacije koje predstavljaju nova znanja koja su dostupna korisnicima. U ovom slučaju je tijekom analize bottom-up. Kada je u pitanju velika količina podataka, tada modeli koji su sposobni aktivno djelovati imaju bitnu ulogu, jer je korisnicima teško formulirati smislene i dobro utemeljene hipoteze.

### **3.4. Primjena rudarenja podataka**

Metode rudarenja podataka mogu se primijeniti na razna područja, od marketinga i proizvodnih procesa do istraživanja rizičnih faktora medicinskih dijagnoza, zatim od procjene učinkovitosti novih lijekova do otkrivanja prijevara.

#### **3.4.1. Relacijski marketing**

Aplikacije rudarenja podataka na području marketinga znatno su pridonijele popularnosti ovih metodologija. Neke od bitnih aplikacija ovog područja su:

- identifikacija segmenata kupaca koji će reagirati na ciljane marketinške kampanje
- identifikacija segmenata kupaca za zadržavanje kampanje
- predviđanje stope pozitivne reakcije na marketinške kampanje
- interpretacija i razumijevanje ponašanja kupaca u trgovini
- analiza proizvoda koje kupac zajedno kupuje prilikom jedne kupovine, poznata kao analiza potrošačke košarice (eng. *market basket analysis*)

#### **3.4.2. Otkrivanje prijevara**

Prijevare se mogu javiti u industrijama kao što su telefonija, osiguranja (lažne potražnje) i bankarstvo (ilegalna uporaba kreditnih kartica i čekova; ilegalne novčane transakcije).

### **3.4.3. Procjena rizika**

Svrha analize rizika je procjena rizika povezanog s budućim odlukama. Primjerice, koristeći dostupna zapažanja iz prošlosti, banka može razviti model za predviđanje da bi utvrdila da li je prikladno dati novčani zajam, ovisno o karakteristikama podnositelja zahtjeva.

### **3.4.4. Rudarenje teksta**

Rudarenje podataka se može primijeniti na različite vrste tekstova, koji predstavljaju nestrukturirane podatke, kako bi se klasificirali članci, knjige, dokumenti, elektronička pošta i web stranice. Najbolji primjeri su web tražilice. Ostale aplikacije rudarenja teksta uključuju filtre elektroničke pošte i interesnih grupa.

### **3.4.5. Prepoznavanje slika**

Obrada i klasifikacija digitalnih slika, statičkih i dinamičkih, pruža velike mogućnosti primjene. Korisno je za raspoznavanje pisanih slova, uspoređivanje i identifikacija ljudskih lica, te otkrivanje sumnjivog ponašanja kamerama za video nadzor.

### **3.4.6. Rudarenje weba**

Koristi se za analize zvane clickstreams, koje prate slijed posjećenih stranica i izbore korisnika Interneta. Mogu se pokazati korisnima za analizu komercijalnih web stranica, pronalaženje najpopularnije stranice ili procjenu učinkovitosti tečaja Internet učenja.

### **3.4.7. Medicinske dijagnoze**

Modeli učenja su neprocjenjiv alat na području medicine, jer omogućuje rano otkrivanje bolesti korištenjem rezultata kliničkih testova.

## **3.5. Prikaz ulaznih podataka**

U većini slučajeva, ulazni podaci za analizu rudarenja podataka imaju oblik dvodimenzionalnih tablica, koje se nazivaju set podataka (eng. *dataset*). Redovi u setu podataka predstavljaju zapise zapažanja iz prošlosti, i nazivaju se primjerima, slučajevima, instancama ili zapisima. Kolone predstavljaju informacije o pojedinom promatranju, i nazivaju se atributima, varijablama, karakteristikama ili značajkama.

Atributi sadržani u setu podataka mogu se svrstati po kategorijama ili numerički, ovisno o vrsti vrijednosti koje poprimaju.

### **3.5.1. Kategorički**

Kategoričke atribute predstavlja konačan broj različitih vrijednosti, koji je obično ograničen na manje od sto, i predstavljaju kvalitativna svojstva entiteta na koje se odnose. Primjerice, atribut županija u kojoj boravi neki pojedinac poprima jednu od vrijednosti iz predodređenog niza, moguće je da ta vrijednost bude cijeli broj. Drugi primjer su telefonske kompanije, prelazak klijenta na mrežu drugog operatera označava se sa vrijednošću 1, dok se ostanak na mreži označava sa 0. Na kategoričke atribute se ne primjenjuju aritmetičke operacije, čak ni kad su kodirane vrijednosti izražene cijelim brojevima.

### **3.5.2. Numerički**

Numeričke atribute predstavljaju konačne ili beskonačne brojčane vrijednosti, nad kojima je moguće obavljati operacije oduzimanja ili dijeljenja. Primjerice, iznos odlaznih telefonskih poziva klijenta, tijekom jednog mjeseca, prikazan je numeričkom varijablom.

Ponekad i preciznija taksonomija atributa može biti korisna.

### **3.5.3. Točke**

Točke su kategorički atributi čija vrijednost može bit točno ili netočno (eng. *true or false*). Ovi atributi mogu biti izraženi koristeći Boolean varijable  $\{true, false\}$  ili binarne varijable  $\{0, 1\}$ . Primjerice, klijent banke može ili ne može koristiti bankovnu kreditnu karticu.

### **3.5.4. Nominalni**

Nominalni atributi su kategorički atributi bez nekog prirodnog slijeda, kao što je županija stanovanja.

### **3.5.5. Redni**

Redni atributi su kategorički atributi koji posuđuju svojstvo prirodnog slijeda, ali nema smisla računati razlike ili omjere između vrijednosti. Primjer za ove atribute je stupanj obrazovanja.



### **3.5.6. Diskretni**

Diskretni atributi su numerički atributi koje predstavljaju konačni brojevi ili brojive beskonačne vrijednosti.

### **3.5.7. Kontinuirani**

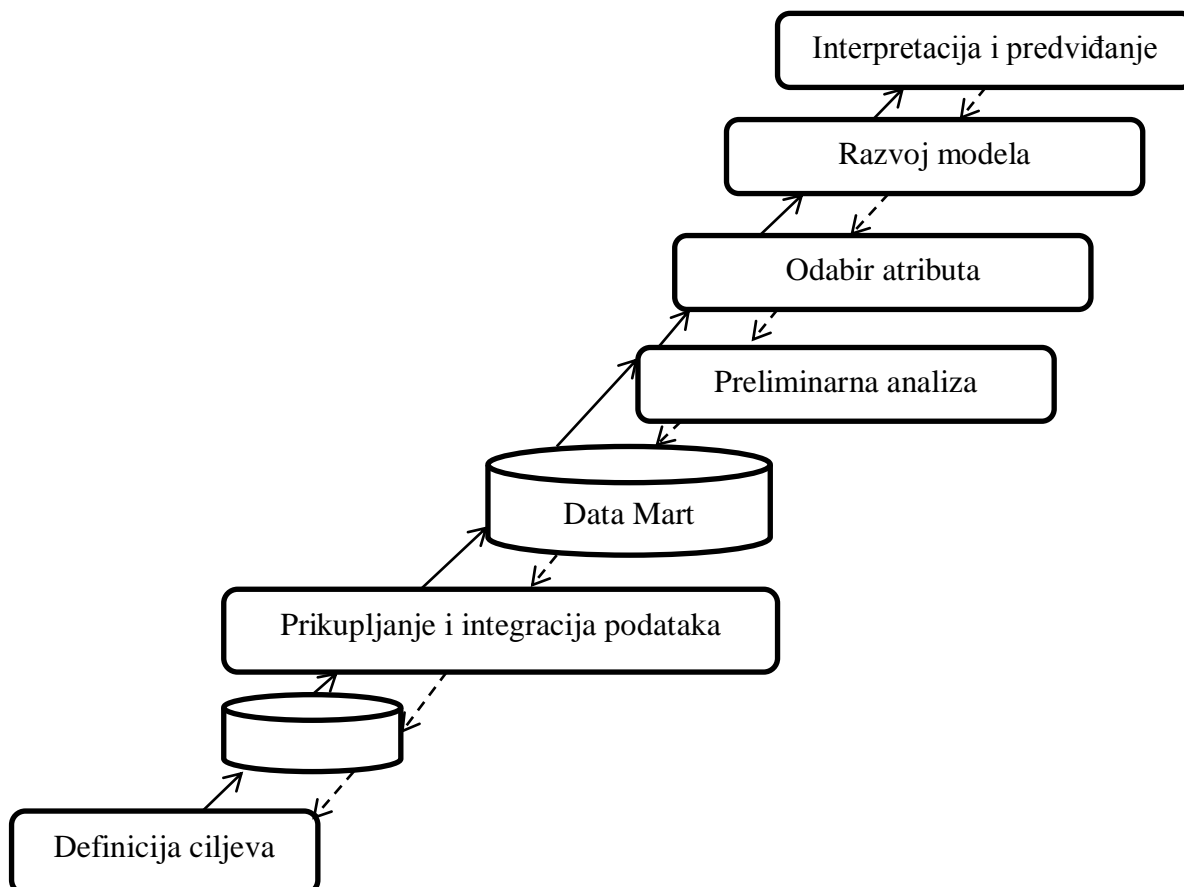
Kontinuirani atributi su numerički atributi koje predstavljaju nebrojive beskonačne vrijednosti.

## **3.6. Proces rudarenja podataka**

Definicija rudarenja podataka koja je dana na početku ovog poglavlja odnosi se na iterativni proces, u kojem modeli učenja i tehnike imaju ključnu ulogu. Slika 3-1 prikazuje glavne faze procesa rudarenja podataka.

### **3.6.1. Definicija ciljeva**

Analize rudarenja podataka provode se određenim stručnim područjima i cilj im je donositeljima odluka pružiti korisno znanje. Zbog toga je potrebna određena sposobnost stručnjaka s tog područja da dobro definira ciljeve istraživanja. Ako problem nije adekvatno identificiran i ograničen, postoji mogućnost da će sve buduće aktivnosti rudarenja podataka biti izvođene uzalud. Ciljevi će biti bolje definirani ako postoji bliska suradnja između stručnjaka na području gdje se primjenjuje analiza i analitičara rudarenja podataka.



Slika 7. Proces rudarenja podataka [1]

### 3.6.2. Prikupljanje i integracija podataka

Podaci mogu potjecati iz različitih izvora, i zato može biti potreba integracija. Izvori podataka mogu biti unutarnji, vanjski ili kombinacija ta dva tipa. Integracija različitih izvora podataka može biti potrebna onda kada podatke treba obogatiti s novim dimenzijama, kao što su varijable s određenog geografskog tržišta, liste potencijalnih kupaca, koje još ne postoje u informacijskom sustavu tvrtke. U nekim slučajevima, izvori podataka su već strukturirani u skladištima podataka ili data marts, i koriste se za OLAP analize i aktivnosti za podršku donošenja odluka. To su vrlo povoljne situacije gdje je dovoljno odabrati attribute koji su potrebni za analizu rudarenja podataka. Međutim, moguće je da su ti podaci iz skladišta podataka agregirani i konsolidirani do te mjere da su beskorisni za bilo kakve analize. Primjerice, ako maloprodajno poduzeće pohranjuje samo ukupan iznos računa, bez da bilježi pojedine artikle, moguće je da će analize rudarenja podataka koje su usmjerene na istraživanje kupovine biti ugrožene.

### 3.6.3. Preliminarna analiza

Preliminarna analiza izvodi se kako bi proučili dostupni podaci i provelo čišćenje podataka. Uglavnom su podaci pohranjeni u skladišta podataka obrađeni za vrijeme učitavanja, na način da su uklonjene sve sintaktičke pogreške. Primjerice, datumi rođenja koji nisu u prihvatljivom razdoblju su uklonjeni, ili negativne naplate u prodaje su ispravljene. U rudarenju podataka čišćenje podataka provodi se na semantičkoj razini. Najprije se proučava raspon vrijednosti svakog atributa, korištenjem histograma za kategoričke attribute i osnovne statistike za numeričke vrijednosti. Na taj način se otkrivaju sve abnormalne vrijednosti i vrijednosti koje nedostaju. Te podatke proučavaju stručnjaci s tog područja kako bi razmotrili isključivanje pojedinog zapisa iz istraživanja.

### 3.6.4. Odabir atributa

Važnost atributa se ocjenjuje ovisno o ciljevima analize. Atributi koji za koje se pokaže da nisu od neke važnosti su uklonjeni, kako bi se seta podataka uklonile nevažne informacije. Zatim slijedi prikladna transformacija i uključivanje novih atributa u set podataka. Primjerice, u većini slučajeva je korisno uvesti nove attribute koji odražavaju trendove podataka, a dobivaju se izračunavanjem omjera i razlika između originalnih varijabli. Preliminarna analiza i odabir atributa su kritične faze procesa rudarenja podataka i mogu uvelike utjecati na uspjeh narednih faza.

### 3.6.5. Razvoj modela

Kad se sastavi kvalitetan set podataka, moguće je početi razvoj modela za raspoznavanje obrazaca i predviđanje. Uglavnom se osposobljavanje (eng. *training*) modela provodi koristeći uzorak zapisa izvađenih iz originalnog seta podataka. Zatim se točnost predviđanja svakog modela procjenjuje pomoću ostatka podataka, koji se dijele na dva dijela. Prvi se sastoji od seta za osposobljavanje (eng. *training set*) i koristi se za identifikaciju specifičnih modela učenja unutar odabrane klase modela. Obično se odabire relativno mala količina uzoraka seta za osposobljavanje, koja je statistički vrlo značajna (nekoliko tisuća uzoraka). Drugi set podataka je test set i koristi se za procjenu točnosti alternativnih modela generiranih tijekom faze osposobljavanja, kako bi se identificirao najbolji model buduća predviđanja.

### 3.6.6. Interpretacija i predviđanje

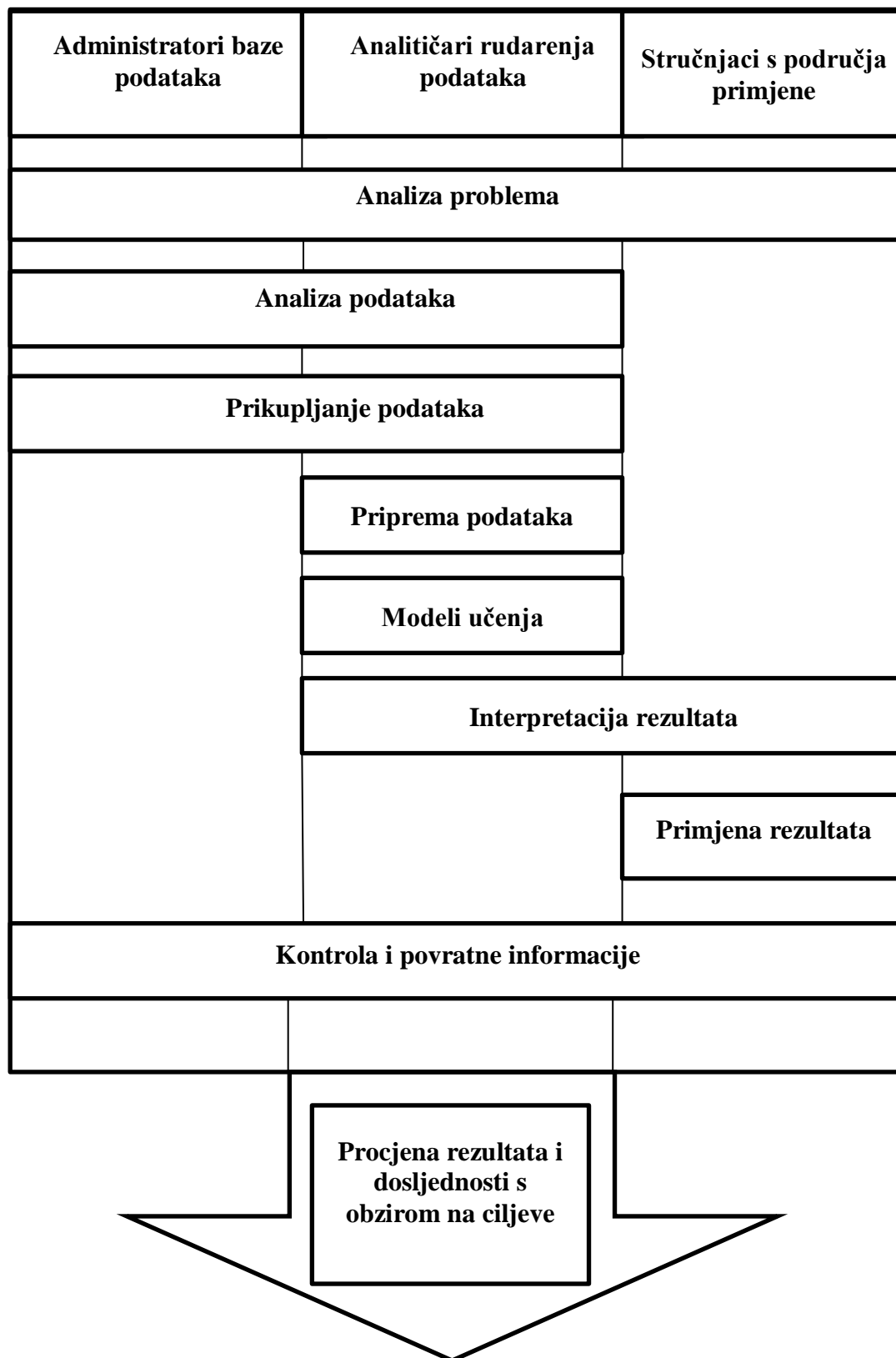
Na kraju procesa rudarenja podataka, koristi se odabrani model, kako bi se postigli zadani ciljevi. Taj se model može primjenjivati i za procedure podrške procesa donošenja odluka, kako bi donositelji odluka imali uvid u predviđanja i stekli veća znanja od proučavanom problemu.

Proces rudarenja podataka uključuje povratne informacije, prikazane isprekidanim linijama na slici 7, koje ukazuju na povratak na prethodne faze ovisno o ishodu naredne.

Treba naglasiti važnosti uključivanja i suradnje više različitih stručnjaka, kako bi se proces rudarenja podataka izveo učinkovito:

- stručnjak područja koje se istražuje, od kojeg se očekuje da definira ciljeve analize, da objasni značenje podataka i da sudjeluje u odabiru najučinkovitijeg i najpreciznijeg modela
- stručnjak za informacijski sustav tvrtke, od kojeg se očekuje da nadzire pristup izvorima informacija
- stručnjak za matematičke teorije učenja i statistiku, koji je potreban za provedbu preliminarne analize i generiranje modela za predviđanje

Slika 8 prikazuje nadležnosti u različitim aktivnostima za svakog sudionika u procesu rudarenja podataka.



Slika 8. Sudionici i njihove uloge u procesu rudarenja podataka[1]

### 3.7. Analiza metodologija

Aktivnosti rudarenja podataka mogu se podijeliti u nekoliko glavnih kategorija, što zavisi o zadacima i ciljevima analize. Ovisno o postojanju tražene varijable, moguće je podijeliti procese učenja na nadzirane (eng. *supervised*) i nenadzirane (eng. *unsupervised*).

#### 3.7.1. Nadzirani procesi učenja

U nadziranoj (direktnoj) analizi učenja ciljani atribut predstavlja klasu kojoj zapis pripada, ili izražava mjerljivu veličinu, kao što je primjerice ukupna vrijednost poziva koje će klijent ostvariti u budućem razdoblju. Nadzirani procesi učenja su orijentirani predviđanju i interpretaciji s obzirom na određeni atribut.

#### 3.7.2. Nenadzirani procesi učenja

Nenadzirani (indirektni) procesi učenja nisu vođeni ciljanim atributom. Zato je, u ovom slučaju, zadatak rudarenja podataka pronaći ponavljajuće obrasce i sklonosti skupa podataka. Primjerice, investicijska kompanija, na temelju transakcija iz prošlosti, želi identificirati skupinu klijenata koji imaju slična investicijska ponašanja. Većini nenadziranih procesa učenja cilj je identificirati skupine zapisa koji su slični unutar svake skupine i različiti od ostalih skupina.

Postoji sedam osnovnih zadataka rudarenja podataka:

- karakterizacija i diskriminacija
- klasifikacija
- regresija
- analiza vremenskih nizova
- asocijativna pravila
- grupiranje
- opis i vizualizacija

Prva četiri zadatka, zadaci su nadziranih analiza rudarenja podataka, koji se koriste ako postoje određene ciljane varijable koje je potrebno obrazložiti na temelju dostupnih atributa ili tijekom razvoja. Ostala tri su zadaci nenadziranih analiza čija je svrha razviti model sposoban izraziti međuodnose dostupnih atributa.

### **3.7.3. Karakterizacija i diskriminacija**

Tamo gdje postoji specifični ciljani atribut, često se prije razvoja klasifikacijskog modela provodi preliminarna analiza, i to zbog dva razloga. Prvi razlog tome je karakterizacija atributa zapisa koji pripadaju istoj klasi, uspoređujući raspodjelu njihovih vrijednosti. Drugi razlog je otkrivanje razlike, uspoređujući raspodjelu vrijednosti atributa zapisa između zadane klase i zapisa drugačije klase, ili između zapisa zadane klase i preostalih zapisa. Ovaj zadatak rudarenja podataka provodi se putem preliminarne analize podataka, i stoga se temelji na upitima koji ne zahtijevaju razvoj specifičnih modela učenja. Informacije koje su prikupljene na ovaj način, korisnicima se prikazuju pomoću histograma i drugih vrsta grafova. Vrijednost generiranih informacija je izvanredna i često se određuje naknadna faza odabira atributa.

### **3.7.4. Klasifikacija**

Kod ovog problema dostupan je skup zapažanja, uglavnom taj skup predstavljaju zapisi seta podataka, čija je ciljana klasa poznata. Primjerice, zapažanja telefonskih kompanija mogu biti binarne klase koje ukazuju da li je određeni klijent aktivan ili je prešao na mrežu konkurencije. Svako zapažanje opisano je određenim brojem atributa čije su vrijednosti poznate. U navedenom primjeru, atributi mogu biti dob, vrijeme koje kupac koristi uslugu i odlazni pozivi. Dakle, algoritmi za klasifikaciju koriste dostupna zapažanja kako bi identificirali model koji može predvidjeti buduću ciljanu klasu čije su vrijednosti atributa poznate. Važno je napomenuti da je ciljani atribut čija se vrijednost predviđa kategoričan u klasifikacijskim problemima i stoga poprima konačan i vrlo mali broj vrijednosti. U većini slučajeva su vrijednosti ciljanih atributa prikazane binarnim varijablama. Kategorička narav ciljanog atributa razlikuje klasifikaciju od regresije.

### **3.7.5. Regresija**

Za razliku od klasifikacije, koja je namijenjena diskretnim vrijednostima, regresija se koristi kada ciljana varijabla poprima kontinuirane vrijednosti. Ovisno o dostupnosti atributa, cilj je predvidjeti vrijednost ciljane varijable za svako zapažanje. Primjerice, ako se želi predvidjeti prodaja proizvoda na temelju promotivnih kampanja i prodajnoj cijeni, ciljana varijabla može poprimiti vrlo visok broj diskretnih vrijednosti te se može tretirati kao kontinuirana varijabla. Klasifikacijski problem se može pretvoriti u regresijski, i obrnuto

### **3.7.6. Analiza vremenskih nizova**

Ponekad ciljani atribut tijekom vremena evoluirá i postane povezan s susjednim vremenskim periodima. U tom slučaju redoslijed vrijednosti ciljane varijable predstavlja vremenski niz. Primjerice, dvije godine se promatra tjedna prodaja određenog proizvoda i predstavlja ju vremenski niz koji sadrži 104 zapažanja. Model za analizu vremenskog niza ispituje podatke, koje karakterizira tijek vremena, i usmjeren je na predviđanje vrijednosti ciljane varijable za jedno ili više budućih razdoblja.

### **3.7.7. Asocijativna pravila**

Asocijativna pravila (grupacije) se koriste za identifikaciju zanimljivih i ponavljajućih asocijacija između grupa zapisa u setu podataka. Primjerice, moguće je odrediti koji se proizvodi kupuju zajedno i koliko često se to događa. Maloprodajne kompanije koriste asocijativna pravila za slaganje proizvoda na policama ili za uređenje kataloga. Grupacije povezanih elemenata također se koriste za kombiniranje proizvoda i usluga.

### **3.7.8. Grupiranje**

Tehnike grupiranja usmjerene su na segmentiranje heterogene populacije u određeni broj podgrupa sastavljenih od opažanja koja dijele slične karakteristike. Za razliku od klasifikacije, grupiranje nema predefinirane klase ili primjere koji ukazuju na ciljanu klasu, tako da su objekti grupirani ovisno o međusobnoj homogenosti. Ponekad identificirane grupe predstavljaju preliminarnu fazu procesa rudarenja podataka, koja se nalazi unutar preliminarne analize podataka. To omogućuje obradu homogenih podataka odgovarajućim tehnikama, te je također moguće smanjiti set originalnih podataka, kako bi se naknadne aktivnosti rudarenja podataka mogle razvijati neovisno za svaku grupu.

### **3.7.9. Opis i vizualizacija**

Ponekad je cilj rudarenja podataka pružiti jednostavan i sažet prikaz podataka pohranjenih u velikom skupu podataka. Iako za razliku od grupiranja i asocijativnih pravila ovaj zadatak ne teži nekoj određenoj grupaciji zapisa u setu podataka, učinkovit i sažet opis informacija vrlo je koristan, jer može ukazati na objašnjena nekih obrazaca i dovesti do boljeg razumijevanja zadanog problema. No nije uvijek lako dobiti smislenu vizualizaciju podataka.



Međutim, trud koji je uložen u vizualni prikaz podataka opravdavaju izvanredne informacije koje su dobivene pomoću dobro osmišljenog grafa.

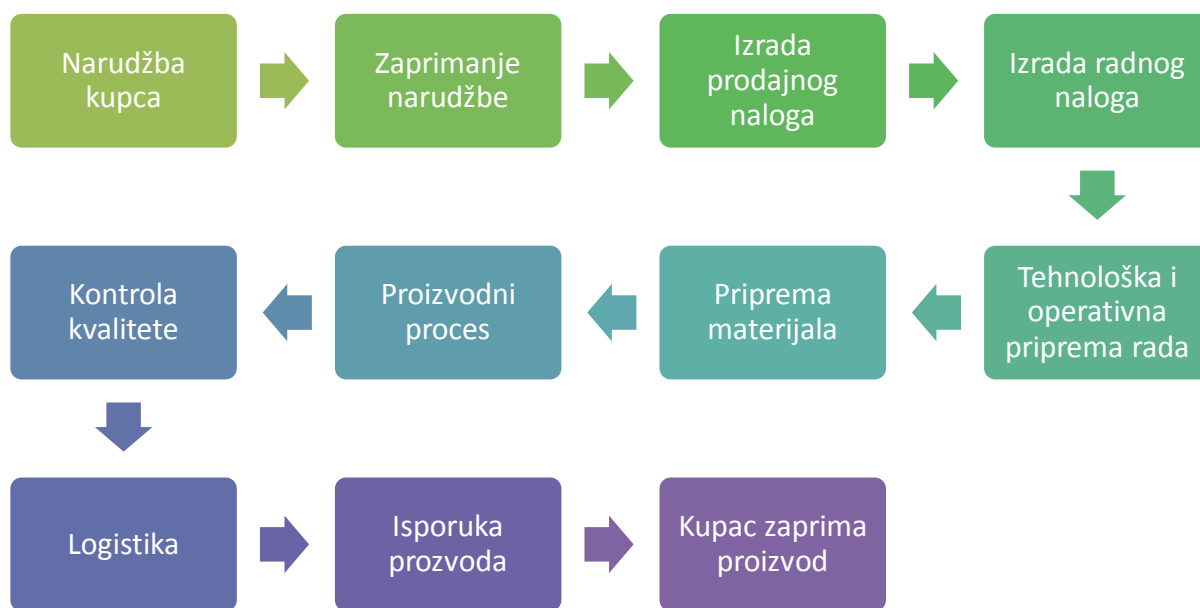
## 4. Primjena sustava poslovne inteligencije u proizvodnji

Organizacije kojima je proizvodnja temeljna djelatnost danas rade u okruženju u kojem se zahtjeva brzo i efikasno donošenje poslovnih odluka. Ove odluke se moraju donositi na ovaj način kako bi organizacija zadovoljila potrebe kupaca. U većini proizvodnih organizacija i poduzeća postoje informacijski sustavi koji prikupljaju podatke o proizvodnji. Ovime se stvara velika količina podataka koju organizacija ne uspijevaju iskoristiti na pravilan način. Ako u poduzeću postoji velika količina podataka bez alata koji omogućava analizu i vađenje informacija iz baze podataka povećava se vrijeme koje zaposlenici moraju provesti za analizu. Ovo zatim dovodi do odgode u procesu donošenja odluka, što zatim dovodi do zastoja u dostavljanju proizvoda kupcu. Kako bi se smanjila taj vremenski razdor, implementiraju se sustavi poslovne inteligencije, kojima je cilj donijeti pravu poslovnu odluku u pravo vrijeme, odnosno biti podrška procesu donošenja odluka u poduzeću. [9]

Proizvodna industrija je jedna od glavnih grana gospodarstva većine zemalja. Kako konkurencija raste, a kupci postaju sve zahtjevniji, poduzeća diljem svijeta su počela tražiti načine kako održati i povećati profit. Jedan od načina je implementacija sustava poslovne inteligencije.

### 4.1. Proizvodni procesi i problemi u proizvodnji

Proizvodnja se sastoji od mnogo procesa koji započinju od narudžbe kupca, a završavaju sa isporukom proizvoda. Općeniti prikaz proizvodnog procesa je prikazan na slici 9. Treba uzeti u obzir da je riječ o općenitom toku procesa u proizvodnji koji može varirati o proizvodu koji se proizvodi.



**Slika 9. Općeniti tok proizvodnje**

U proizvodnom dijelu nailazi se na problem kada zaposlenici nemaju sustav pomoću kojeg se omogućava praćenje proizvodnje u realnom vremenu. Kada bi u tom slučaju htjeli provjeriti stanje proizvoda morali bi otići u odjel proizvodnje i tamo provjeravati proizvod. Problem koji se ovdje može javiti je kada kupac ima određene zahtjeve za neki proizvod, a zbog nemogućnosti provjeravanja svakog proizvoda ne provjere se njegovi zahtjevi. Zaposlenici u proizvodnji će često prvo odabirati proizvod koji je jednostavniji za izradu, što uzrokuje stvaranje vremenskog razdora između narudžbe i isporuke proizvoda. [5, 6]

U proizvodnji se također javlja i problem nestrukturiranosti podataka. Često u poduzećima postoje informacijski sustavi koji služe za praćenje proizvodnje, i koji skupljaju veliku količinu podataka. Podaci se skupljaju iz svih dijelova proizvodnje, poput montaže, planiranje resursa i materijala, financija, rasporeda radnika i strojeva itd. Problem je iskoristiti te podatke na način koji će dati razumljive i korisne informacije i znanja.

ERP (eng. *Enterprise Resource Planning*) sustavi služe za planiranje materijalnih potreba, a također mogu poslužiti i za povezivanje drugih dijelova poduzeća. Problem koji se javlja kod ERP sustava je što ne uspijeva iskoristiti prikupljene podatke. Podaci prikupljeni ERP sustavom ne mogu se iskoristiti za detaljnije analize za koje su pogodni sustavi poslovne

inteligencije, poput analize vremenskih nizova ili rudarenja podataka. U takvim prilikama vide se prednosti sustava poslovne inteligencije. Sustavi poslovne inteligencije mogu pretvoriti veliku količinu podataka u strukturirane podatke koji se kasnije pretvaraju u informacije i znanja koji služe za podršku procesu donošenja odluka. [9]

Sustavi poslovne inteligencije pomoću tako strukturiranih podataka iz raznih izvještaja i tablica mogu izvući vrijedne informacije, predvidjeti buduće narudžbe, pa se tako mogu i bolje isplanirati materijalne potrebe poduzeća.

## **4.2. Uvođenje sustava poslovne inteligencije u proizvodne organizacije**

Proizvodna poduzeća su pred sve većim pritiskom zbog visokog stupnja konkurencije, kao i zbog sve većeg razvoja globalnog tržišta. Ovi problemi se javljaju u vremenima slabije ekonomske moći raznih organizacija, čime se povećava potreba za povećanjem prihoda, smanjenjem troškova i povećanjem mogućnosti proizvodnje. Problemi se također javljaju i uslijed sve većih zahtjeva tržišta. Od proizvodnih poduzeća se tako očekuje proizvodnja velikog broja različitih proizvoda, od kojih svi moraju zadovoljiti sve standarde proizvodnje. [8]

Postavlja se pitanje kako proizvodna poduzeća mogu zadovoljiti navedene uvjete, a pritom osigurati povećanje prihoda i profita ili čak prihvaćanje novih poslovnih prilika.

### **4.2.1. Sustavi poslovne i proizvodne inteligencije**

Kada je riječ o proizvodnim poduzećima, postoje dvije vrste inteligencije. To su poslovna i proizvodna inteligencija. Važno je znati razliku kako bi znali primijeniti odgovarajući alat za analizu. Poslovna i proizvodna inteligencija su usko povezane, a razlika se očituje u tome što je poslovna inteligencija sveobuhvatna. Sustavi poslovne inteligencije tako obuhvaćaju sve faktore koji ulaze u analizu, od vremenske prognoze pa sve do analize ponašanja kupaca u određenim razdobljima.

Za razliku od sustava poslovne inteligencije, sustavi proizvodne inteligencije su fokusirani na proizvodni proces, i to od ideje za neki proizvod pa sve do kraja životnog vijeka proizvoda. Ovakve analize služe i za planiranje recikliranja proizvoda, ali se najviše fokusiraju na sami proizvodni proces. Sustavi proizvodne inteligencije provede analize u realnom vremenu, i za to su potrebne informacije u realnom vremenu o svim događanjima na

odjelu proizvodnje. Kada sustav mjeri, prati, analizira i upozorava na probleme u odjelu proizvodnje, riječ je o sustavima proizvodne inteligencije.

#### **4.2.2. Značaj podataka u proizvodnji**

U proizvodnim procesima želimo znati kako će se određeni materijali ponašati u odrađenom vremenu, ili kako će neki proizvodni proces utjecati na lanac opskrbe. Ovdje se javlja problem sakupljanja podataka sa svih razina proizvodnje. Trenutak kada poduzeće započne integraciju podataka sa lancem opskrbe, počinje se stvarati velika količina podataka koja uključuje više izvora podataka, čime se otežava praćenje i mjerenje efikasnosti proizvodnje.

Ideja proizvodne inteligencije je povezivanje svih ovih tipova podataka sa podacima iz prodaje. Poduzeća imaju neke proizvode koje žele prodati, pa se postavlja pitanje kako menadžment poduzeća može donijeti najbolju odluku ako nema sveobuhvatni i strukturirani prikaz proizvodnje i prodaje. Uvođenjem sustava poslovne inteligencije u proizvodno poduzeće omogućava se optimalni menadžment proizvodnih stanica. Velika proizvodna poduzeća imaju preko 90 proizvodnih stanica pa je očito da žele optimalnu iskoristivost svake stanice. Ukoliko poduzeće uspije u optimalnom menadžmentu svih proizvodnih stanica stvara se prednost nad konkurencijom.

Za siguran povrat sredstava uvođenjem sustava poslovne inteligencije preporuča se da poduzeće posjeduje barem 3-5 tvornica prosječne veličine. Manja poduzeća svejedno mogu izvući korist iz fokusiranih sustava proizvodne inteligencije koji su jeftiniji za implementaciju. [10]

#### **4.2.3. Prednosti sustava poslovne inteligencije u proizvodnji**

Uvođenjem sustava poslovne inteligencije proizvodna poduzeća ostvaruju niz prednosti. Menadžmentu se omogućava razumijevanje rizika i težine proizvodnje nekog proizvoda. Na temelju analiza sustava poslovne inteligencije menadžment poduzeća može prihvatiti ili odbiti proizvodnju nekog proizvoda koji se na prvi pogled čini iznimno profitabilnim, ali u stvarnosti bi bila pogreška pristati na proizvodnju.

Sustavi poslovne inteligencije u proizvodnji predstavljaju sve više i više sigurna ulaganja za razna poduzeća. Menadžment poduzeća počinje shvaćati koliko je velika razlika

između proizvodne inteligencije koja omogućava nesmetanu i sigurnu proizvodnju i sustava poslovne inteligencije. Poduzeća koja počinju sa implementacijom sustava poslovne inteligencije žele odgovore na mnoga pitanja, kao na primjer, gdje početi, kada će se ostvariti povrat sredstava, tko su ljudi koji trebaju biti uključeni u projekt te tko bi trebao biti glavni i odgovorni za uspjeh projekta sustava poslovne inteligencije?

Odgovorna osoba mora biti u potpunosti povezana sa projektom razvoja sustava poslovne inteligencije. U razvoju ovakvih sustava vidi se značaj povezanosti više odjela u poduzeću. Ukoliko osoba iz odjela menadžmenta treba neku informaciju, u organizaciji bez sustava poslovne inteligencije doći će do osobe iz odjela informacijskih sustava. Nakon toga osoba zadužena za povrat informacije prosljeđuje informacije menadžmentu. Sa sustavom poslovne inteligencije menadžment može dobiti točno određeni tip informacija u realnom vremenu.

Kako bi projekt sustava poslovne inteligencije bio što uspješniji, osoba zadužena za razvoj sustava mora pronaći slabe točke poduzeća i brzo ih ukloniti tijekom implementacije. U isto vrijeme, prevelike promjene tijekom implementacije mogu prouzrokovati veliko povećanje troškova. [10]

#### **4.2.4. Definiranje ciljeva projekta**

Čak i ukoliko poduzeće nema nekih slabih točaka ili ako te slabe točke ne opravdavaju implementaciju sustava poslovne inteligencije, to ne znači da poduzeće implementacijom sustava poslovne inteligencije ne može izvući veliku korist. Pametna poduzeća žele provjeriti postoje li informacije koje mogu pomoći u procesu donošenja poslovnih odluka kako bi se ispunili ciljevi poduzeća. Zato je iznimno važno da poduzeće jasno definira ciljeve prilikom implementacije sustava poslovne inteligencije. Prilikom razvoja sustava poslovne inteligencije od velike je važnosti dobra komunikacija i povezanost troje odjela poduzeća:

- Eksperti u informacijskim sustavima
- Eksperti u proizvodnji
- Menadžment poduzeća

U slučaju dobre komunikacije i suradnje ova tri odjela povećava se vjerojatnost uspjeha projekta razvoja sustava poslovne inteligencije.

#### 4.2.5. Implementacija sustava poslovne inteligencije

Implementacija sustava poslovne inteligencije predstavlja veliki izazov i za najveća poduzeća. Poduzeća koja imaju svoje podružnice u različitim državama imaju velikih problema sa generiranom velikom količinom podataka. Važno je su podaci strukturirani kako bi ih mogli iskoristiti za analizu pomoću alata poslovne inteligencije. Često je problem upravo nestrukturiranost podataka iz više izvora (eng. *data sources*). [6]

##### 4.2.5.1. Definiranje podataka

Menadžment poduzeća mora razumjeti koliko je bitno definiranje podataka. Ukoliko podaci nisu dobro definirani, neće biti ni dobrih rezultata u analizi. Poduzeće mora razumjeti koliki je prihod ostvaren s obzirom na razne faktore. Primjerice, lagano je odrediti koliko koji kupac donosi prihoda poduzeću, ali poduzeće se mora zapitati koliko taj kupac zapravo košta poduzeće? Moguće je odrediti trošak materijala, ali potrebno je odrediti i troškove radne snage, utrošene električne energije, a to nisu podaci koji su uvijek isti za sve kupce.

##### 4.2.5.2. Integracija podataka

Sustavi poslovne inteligencije se često isporučuju za svojim skladištem podataka. Kako alati poslovne inteligencije postaju sve bolji i bolji u prikupljanju podataka iz više izvora, tako je sve manje i manje izražena potreba za klasičnim skladištima podataka. U poduzeću mora postojati neko logičko mjesto, koje ne mora nužno biti fizičko, na kojem su svi podaci iz više izvora skupljeni, kako bi bilo jednostavnije povezati i usporediti sve podatke, te zatim i provesti analize pomoću alata poslovne inteligencije.

##### 4.2.5.3. Pročišćavanje podataka

Proizvodno orijentirana poduzeća često generiraju veliku količinu podataka koja nije uvijek korisna. Zbog toga mora postojati plan upravljanja velikom količinom podataka koji će osigurati da su podaci u analizi sustava poslovne inteligencije točni i precizni. Sinkronizacija, integracija, spajanje i pročišćavanje baza podataka je veliki izazov i za najveće kompanije, jer nakon su provedeni ovi procesi, baza podataka je spremna za analizu, nakon čega poduzeće dobiva informacije i znanja od velike važnosti, a time ostvaruje i konkurentsku prednost.

### **4.3. Poslovna inteligencija u proizvodnji**

Poslovna inteligencija je definirana kao skup metoda pomoću kojih se podaci pretvaraju u informacije, a zatim u znanje. Sustavi poslovne inteligencije, kada su u potpunosti integrirani u poduzeću, omogućavanje zaposlenicima uvid u proizvodne procese u realnom vremenu. Analizom podataka se kasnije upravlja poslovnim procesima, poboljšava se kvaliteta rada te se otkrivaju nove prilike na globalnom tržištu. Danas se sustavi poslovne inteligencije koriste u proizvodnji kako bi se riješili problemi u organizaciji poduzeća iz poslovne perspektive, a posebno se koriste kao podrška procesu donošenja odluka kako bi poduzeće održalo svoju poziciju na tržištu. U nastavku rada biti će ukratko prikazane neke grane proizvodnje i mogućnosti poboljšavanja koje nude sustavi poslovne inteligencije.

#### **4.3.1. Proizvodnja poluvodiča**

Prilikom proizvodnje poluvodiča javljaju su se sljedeći problemi:

- Narudžbe su se sortirale samo po datumu
- Pretpostavili su se neograničeni kapaciteti
- Plan proizvodnje je bio vremenski jako zahtjevan

Pomoću sustava poslovne inteligencije uveo se sustav za podršku procesa donošenja odluka. Nakon uvedenog sustava poslovne inteligencije narudžbe kupaca sortirale su se u realnom vremenu, a datumi isporuke su se određivali na temelju kapaciteta, čime se povećala efikasnost i pouzdanost. [9]

#### **4.3.2. Proizvodnja cementa**

Promatrani su problemi koji se javljaju prilikom proizvodnje cementa, i to u sektoru financija i proizvodnje.

Problemi koji su uočeni uključivali su nekvalitetno dijeljenje informacija koje je dovelo do nekonzistentnosti i nepreciznosti informacija. Predloženo rješenje pomoću sustava poslovne inteligencije je bila izrada web stranice preko koje su se podaci unosili u realnom vremenu. Rezultati su bili smanjenje troškova za 3% i povećanje efikasnosti u proizvodnji za 5%. [9]



### **4.3.3. *Proizvodnja kemikalija***

U kemijskoj industriji javljaju se razni problemi za koje su pogodni sustavi poslovne inteligencije. Najveći problem bio je manjak komunikacije između raznih odjela, što je stvaralo zastoje u procesu donošenja odluka. Još jedan problem koji se javlja u ovoj industriji je smanjenja proizvodnosti, što je posljedica nemogućnosti odgovora u realnom vremenu na probleme u proizvodnji. Također nije postojala komunikacija između plana proizvodnje i stvarnoga proizvodnog odjela, što je predstavljalo visok trošak.

Uvođenjem sustava poslovne inteligencije ostvarile su se uštede u proizvodnom procesu, i to od 3 do 5%. Također je došlo i do povećanja proizvodnje između 8 i 10%. Uvođenjem ovakvog sustava povećala se i povratna informacija kupaca. [9]

### **4.3.4. *Proizvodnja elektronskih komponenata***

Cilj je provjeriti kako sustav poslovne inteligencije može poboljšati poslovanje poduzeća koje proizvodi elektronske komponente. Uočeni su problemi u proizvodnji, skladištenju i prodaji. Problemi koje je bilo potrebno riješiti bili su predviđanje potreba globalnog tržišta, loš servis i loši sustav stvaranja izvještaja. Nakon uvođenja sustava poslovne inteligencije profit poduzeća povećao se za 36%, a sve informacije bile su dostupne u realnom vremenu. Povezali su se razni odjeli poduzeća, poput prodaje, planiranja i predviđanja, proizvodnje, analize profita i izvještavanja u realnom vremenu. [9]

## 5. Rudarenje podataka pomoću programskog sustava RapidMiner

U ovom poglavlju biti će opisane osnove rada u programskom paketu RapidMiner, koji omogućava napredno rudarenje podataka pomoću ranije opisanih metoda. U praktičnom dijelu rada biti će prikazana mogućnost korištenja integriranih alata poslovne inteligencije u programskom paketu RapidMiner na bazi podataka iz jednog poduzeća. Cilj u ovoj bazi podataka biti će korištenje što je više moguće alata poslovne inteligencije kako bi se dokazalo da oni služe kao potpora procesu donošenja odluka. Opisane su metode i znanja potrebna za predprocesuiranje baze podataka i prilagodbe podataka određenim metodama rudarenja podataka. Na bazi podataka proizvodnog poduzeća će se pomoću algoritma za kreiranje asocijativnih pravila stvoriti pravila pomoću kojih će se moći pronaći pozicije koje se nalaze na istim radnih nalogima. Pomoću ove analize se budućim kupcima može predložiti koji proizvodi su im potrebni ukoliko naručuju jednu vrstu proizvoda. Također su kreirana i dva modela pomoću neuronskih mreža i alata za kreiranje poslovnih pravila. Ovo su prediktivni modeli, dok je model za kreiranje asocijativnih pravila opisni.

### 5.1. Opis baze podataka

U bazi podataka postoje razna polja koja predstavljaju značajke proizvodnje. Potrebno je napraviti analizu i predprocesuiranje baze podataka kako bi mogli primijeniti alate poslovne inteligencije. U bazi podataka u Excel formatu postoji 4 stranica, a nakon predprocesuiranja 5. Predprocesuiranje je bilo potrebno napraviti prilikom kreiranja modela za stvaranje asocijativnih pravila. Predprocesuiranje će biti objašnjeno u sljedećim poglavljima. U tablicama su dani opisi podataka za svaku stranicu u MS Excel bazi podataka. Vidimo kako na više radnih stranica postoje ista polja. Predprocesuiranje baze podatak želimo postići čistu bazu, odnosno bazu podatka koja je prigodna za kreiranje modela pomoću programskog sustava RapidMiner. Podaci na koje je potrebno obratiti pažnju jesu numeričkog tipa, a to su lansirana količina proizvoda, broj radnog naloga, pozicija, suma minuta, suma operacija, faktor cijene, broj podnaloga i vrijeme do isporuke. Ove podatke možemo iskoristiti u prediktivnom modelu pomoću neuronskih mreža. Za ostale podatke potreban nam je algoritam za induktivno učenje. Za kreiranje asocijativnih pravila moguće je koristiti i numeričke i tekstualne tipove podataka ako je baza podataka pravilno predprocesuirana.

Na radnom listu NAL.PODNAL.TEHNO postoje sljedeća polja:

**Tablica 2. Opis polja na radnom listu NAL.PODNAL.TEHNO**

STDN	Broj radnog naloga
LANSKOL	Količina pozicija lansiranih u proizvodnju
EL	Oznaka standardne pozicije
POZ	Broj pozicije radnog naloga
NAZIV	Naziv standardne pozicije
SumaMinuta	Ukupne planirane minute za izradu pozicije (sve operacije)

Na radnom listu NAL.PODNAL.TEHNO.DATUM nalaze se sljedeća polja:

**Tablica 3. Radni list NAL.PODNAL.TEHNO.DATUM.**

DatumNarudzbe	
DatumIsporuke	
TjedanNarudzba	Tjedan u godini
MjesecNarudzba	Mjesec u godini
TjedanIsporuka	Tjedan u godini
MjesecIsporuka	Mjesec u godini
VrijemeDoIsporuke	DatumIsporuke- DatumNarudzbe
DrzavaIsporuke	
FaktorCijene	
Nalog	Vidi STDN
TipNalog	Tip naloga po standardu
BrojPodnaloga	Broj različitih vrsta pozicija na nalogu
SumaMin	Ukupne planirane minute za izradu naloga (sve pozicije i operacije)

Na radnom listu PODNAL.TEHNO nalaze se sljedeća polja:

**Tablica 4. Radni list PODNAL.TEHNO**

EL	Oznaka standardne pozicije
NAZIV	Naziv standardne pozicije
SumaMinuta	Ukupne planirano zauzeće resursa u minutama

Na radnom listu PODNAL.TEHNO.TIPRM nalaze se sljedeća polja:

**Tablica 5. Radni list PODNAL.TEHNO.TIPRM**

EL	Oznaka standardne pozicije
NAZIV	Naziv standardne pozicije
SumaMinuta	Ukupne planirano zauzeće resursa u minutama
TipObrade	Grupa strojeva na kojima je planirano odvijanje operacija

## 5.2. Model za pronalaženje asocijativnih pravila

Izrada modela za pronalaženje asocijativnih pravila zahtjeva bazu podataka koja je predprocesuirana, kako bi podaci u bazi bili u točno određenom obliku koji zadovoljava potrebe algoritma. Nakon predprocesuiranja je moguće bazu podataka učitati u programski paket RapidMiner i zatim pomoću operatora u RapidMiner-u provesti algoritam pomoću kojega se vide povezane pozicije na različitim nalogima.

### 5.2.1. Predprocesuiranje baze podataka za izradu modela asocijativnih pravila

U bazi podataka koju je potrebno obraditi nalazi se više stavki, od kojih nam za asocijativna pravila većina nije bitna. Za pronalaženje asocijativnih pravila važne su stavke NAZIV i NALOG. Ova dva polja odabrana su jer pomoću njih je moguće otkriti koje su pozicije povezane i ovise jedna o drugoj. U željenoj bazi podataka stavka NALOG postaje ID, odnosno u slučaju trgovačkog centra broj izdanog računa. Pomoću stavke NAZIV sada trebamo dobiti bazu podataka u kojoj se vidi koje su stavke na kojem nalogu. Ovdje se nailazi na problem koji nije bilo moguće riješiti pomoću programskog paketa RapidMiner. U nestrukturiranoj bazi podataka je svaka pozicija posebno imala svoj ID, tako da su se u više slučajeva pojavljivali isti nalozi, odnosno isti ID. Predprocesuiranje je zbog ovog problema bilo potrebno napraviti u programskom paketu MS Excel. Na slici 10 je ilustrirana nestrukturirana baza podataka.

H	I	J	K	L	M	N
Drzavalsporuke	VrijemeDolsporuke	DatumNarudzbe	DatumIsporuke	SumaMinutaPodnalog	SumaOperacijaPodnalog	Nalog
Ruska Federacija	32	5/26/2014	6/27/2014	2339		15 14-0380
Velika Britanija	27	2/5/2014	3/4/2014	296		9 14-0086
Velika Britanija	18	1/17/2014	2/4/2014	3181		27 14-0025
Italija	69	4/8/2014	6/16/2014	4172.4		32 14-0271
Njemačka	39	4/5/2014	5/14/2014	1115		14 14-0259
Njemačka	66	6/17/2014	8/22/2014	526		11 14-0453
Velika Britanija	30	5/21/2014	6/20/2014	1122		11 14-0361
Poljska	24	3/18/2014	4/11/2014	2829		15 14-0185
Češka	29	6/25/2014	7/24/2014	1624		17 14-0478
Češka	26	7/2/2014	7/28/2014	4189		13 14-0498
Češka	42	6/12/2014	7/24/2014	799		15 14-0440
Ruska Federacija	35	4/8/2014	5/13/2014	840		16 14-0272

Slika 10. Prikaz nestrukturirane baze podataka

Iz ovakve baze podataka nije moguće provesti algoritam za kreiranje asocijativnih pravila. Bazu je potrebno predprocesuirati. Postupak predprocesuiranja je proveden u programskim paketima MS Excel i RapidMiner. Prvo je potrebno izolirati potrebne podatke, odnosno ostaviti samo podatke potrebne za analizu.

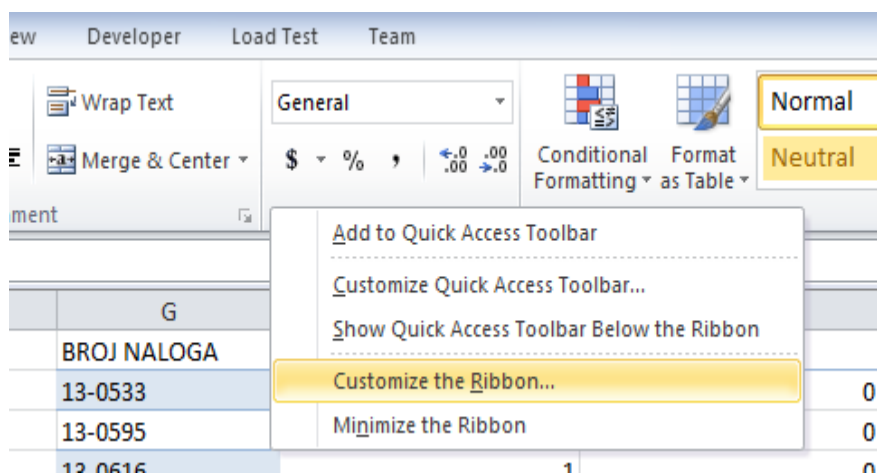
### 5.2.1.1. Predprocesuiranje u programskom paketu MS Excel

U programskom paketu MS Excel je prvo provedena analiza podataka. Zaključeno je kako su nam podaci koji su potrebni za daljnju analizu u kolonama NALOG i NAZIV. Ostale kolone za ovo analizu nisu potrebne te se izbacuju iz tablice.

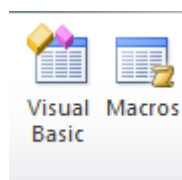
	A	B
1	NAZIV	Nalog
2	PRSTEN	13-0533
3	DNO PRETKALUPA	13-0533
4	ČAHURA ZA KAP	13-0533
5	GLAVA ZA PUHANJE	13-0533
6	JEZGRENIK	13-0533
7	DNO KALUPA	13-0533
8	JEZGRENIK	13-0533
9	CANGA	13-0533
10	KALIBAR	13-0533
11	PRETKALUP	13-0533
12	KALUP	13-0533
13	ČAHURA ZA KAP	13-0595
14	DNO PRETKALUPA	13-0595
15	PRETKALUP	13-0595
16	KALUP	13-0595
17	PLOČICA KALUPA	13-0595
18	PLOČICA KALUPA	13-0595
19	DNO KALUPA	13-0595

**Slika 11. Sortirana baza podataka**

Sa slike 11 vidi se kako izgleda sortirana baza podataka. Također se vidi i sljedeći problem. Za jedan nalog postoji više pozicija, a svaka je u posebnom retku. Ovaj problem nije bilo moguće riješiti u programskom paketu RapidMiner te se pristupilo programiranju u dodatku MS Excel-a. U programskom paketu MS Excel je implementiran dodatak Visual Basic for Applications, koji je ujedno i njegov programski jezik. Pomoću ovog dodatka je moguće kreirati Macro modul, koji služi za automatsko popunjavanje tablice, ali pod određenim uvjetima. Prvo je potrebno omogućiti ovaj dodatak. Na početnom ekranu MS Excel-a je potrebno na alatnoj traci kliknuti desnu tipku miša i odabrati opciju „Customize the Ribbon“ kako je prikazano na slici 11. Nakon toga je potrebno uključiti opciju „Developer“, čime se omogućava kreiranje i pisanje programskog koda koji će biti potreban za pravilno popunjavanje baze podataka. Nakon toga pristupa se Visual Basic programskom jeziku.

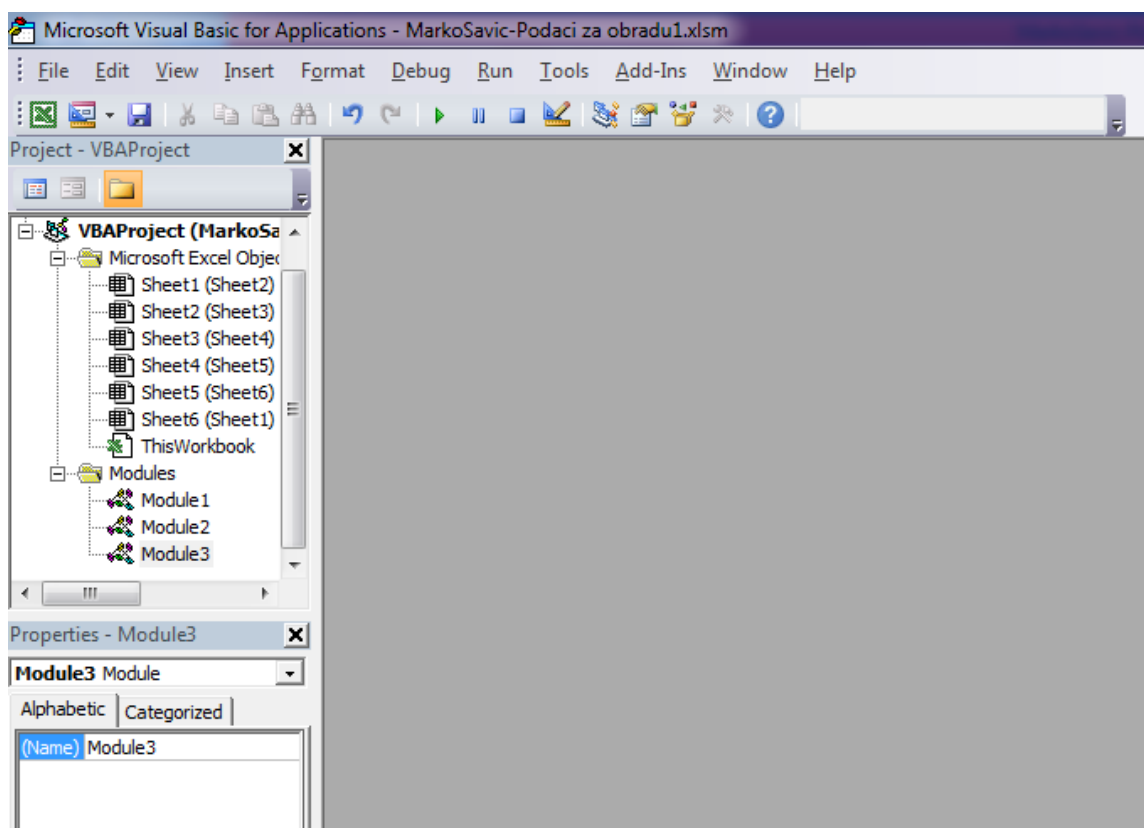


**Slika 12. Podešavanje alatne trake MS Excel-a**



**Slika 13. Pristupanje Visual Basic programskom jeziku**

Nakon što je uključen Visual Basic programski jezik, potrebno je kreirati novi modul i spremiti ga kao Macro. Za ovaj primjer korištena su dva Macro modula, jedan je popunjavao tablicu, a drugi je izbrisao sve podatke, kako bi se mogla provjeriti ispravnost unesenih vrijednosti. Kako bi predprocesuiranje u RapidMiner programskom paketu bilo jednostavnije, korištene su vrijednosti 0 i 1, čime RapidMiner automatski radi razliku za binominalne vrijednosti. Na slici 14 prikazano je početno sučelje Visual Basic for Applications dodatka.



**Slika 14. Sučelje Visual Basic dodatka**

Sada je potrebno kreirati novi modul, a unutar njega dva spomenuta Macro modula. Na slikama 15 i 16 je prikazan kod koji je potrebno unijeti u modul Visual Basic dodatka kako bi se kreirala dva Macro modula koji popunjavaju i prazne tablicu prema prije opisanim zahtjevima.

```
(General)
Sub Sort()
    Dim i As Integer
    Dim j As Integer
    Dim k As Integer
    Dim c As Integer

    j = 2
    k = 2

    While Worksheets("Sheet6").Cells(k, 20).Value <> Blank

        While Worksheets("Sheet6").Cells(j, 14) = Worksheets("Sheet6").Cells(k, 20)

            For i = 21 To 68

                If Worksheets("Sheet6").Cells(1, i) = Worksheets("Sheet6").Cells(j, 1) Then

                    Worksheets("Sheet6").Cells(k, i).Value = 1

                End If

            Next i

            j = j + 1

        Wend

        For i = 21 To 68

            If Worksheets("Sheet6").Cells(k, i).Value <> 1 Then

                Worksheets("Sheet6").Cells(k, i).Value = 0

            End If

        Next i

        k = k + 1

    Wend
```

**Slika 15. Programski kod Macro modula Sort**

Macro modul „Sort“ služi za popunjavanje tablice vrijednostima 0 i 1, a radi tako da uspoređuje vrijednosti u prvom retku tablice sa brojem naloga i nazivom. Brojevi naloga se također moraju poklapati kako bi bio zadovoljen uvjet potreban za upisivanje vrijednosti „1“ u tablicu. Na ovaj način popunjavamo bazu podataka tako da bude spremna za analizu i kreiranje asocijativnih pravila u RapidMiner-u.



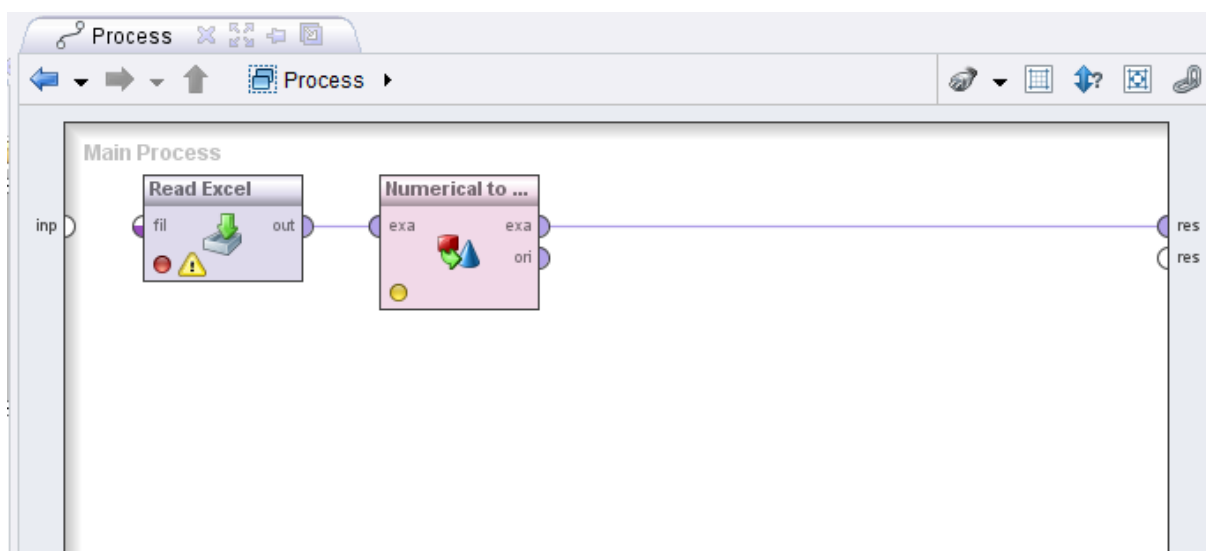
```
⇒ Sub prazn()  
    Dim k As Integer  
    Dim i As Integer  
    For k = 2 To 573  
        For i = 21 To 68  
            Worksheets("Sheet6").Cells(k, i).Value = Blank  
        Next i  
    Next k  
End Sub
```

**Slika 16. Programski kod Macro modula prazn**

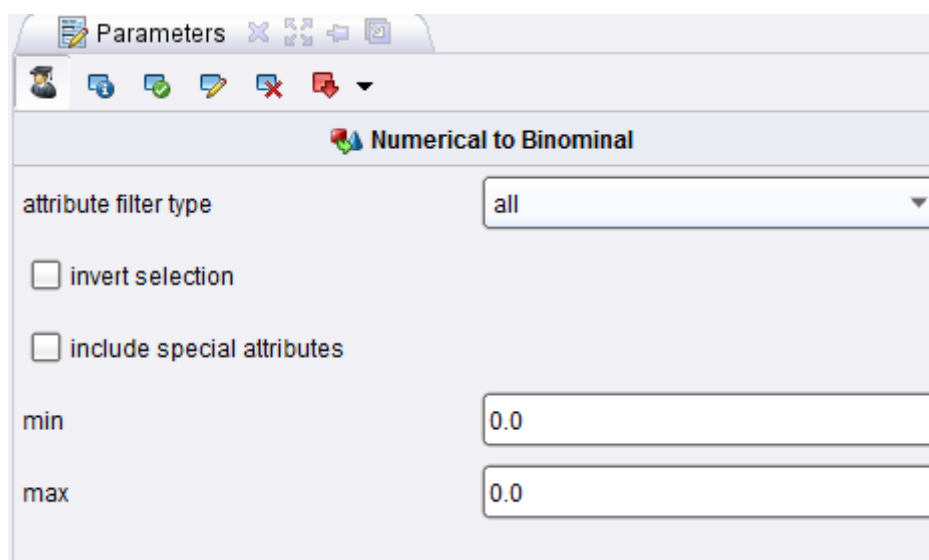
Macro modul prazn služi za jednostavno brisanje svih vrijednosti iz baze podataka. Ovaj Macro modul smo koristili kao jednostavnu mogućnost brisanja i provjere upisanih vrijednosti. Ovaj modul jednostavno bira radnu stranicu sa koje želimo obrisati vrijednosti iz željenih redova i kolona.

#### 5.2.1.2. *Predprocesuiranje u programskom paketu RapidMiner*

U programskom paketu RapidMiner ostalo je samo predprocesuiranje baze podataka kako bi podaci bili u traženom obliku. U slučaju kreiranja asocijativnih pravila pomoću operatora FP rasta i algoritma za kreiranje asocijativnih pravila potrebni su nam podaci u binominal obliku. U programskom paketu RapidMiner za ovaj problem koristimo operater „Numerical To Binominal“, koji odabrane numeričke vrijednosti atributa pretvara u binominal oblik. Binominal oblik je „true-false“ oblik, gdje vrijednost 1 predstavlja true, a 0 false. Ovakav oblik nam je potreban za kreiranje frekventnih podskupova pomoću algoritma FP rasta. Na slici 18 prikazan je model pomoću kojega pretvaramo numeričke vrijednosti u binominal oblik. Također, na slici 19 su prikazane postavke operatora „Numerical To Binominal“, odnosno granica između vrijednosti koje operator pretvara u „true-false“ oblik.



Slika 17. Model za pretvorbu numeričkih vrijednosti



Slika 18. Postavke „Numerical to Binominal“ operatora

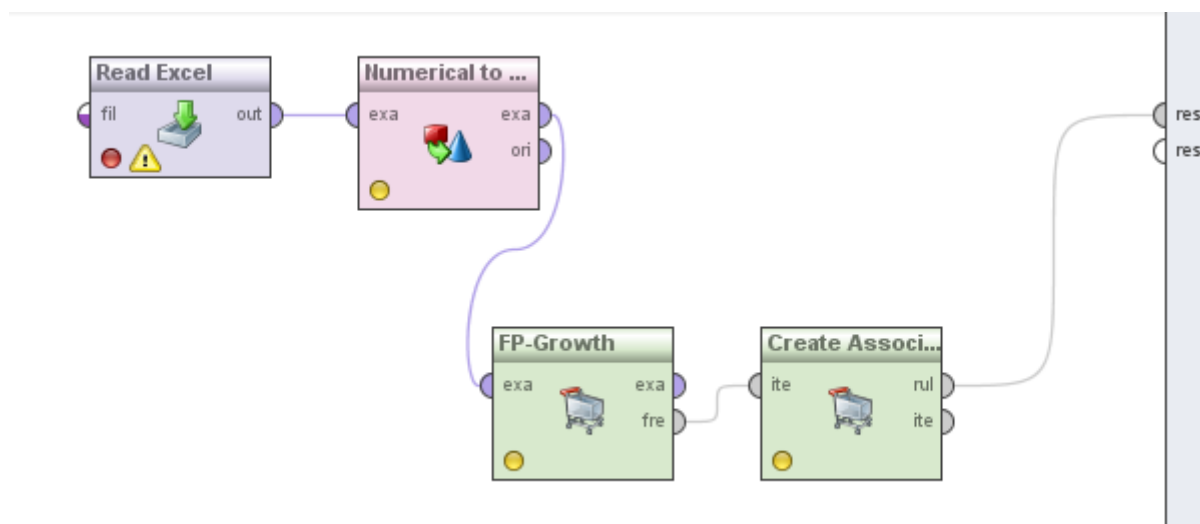
Prema početnim postavkama operatora „Numerical to Binominal“ u programskom paketu RapidMiner postavljena je vrijednost 0 i za gornju i za donju granicu za „true-false“ formu. Ovakav način odgovara problemu zbog toga što će RapidMiner u tom slučaju sve numeričke vrijednosti koje iznose 0 pretvoriti u „false“ oblik, a sve ostale u „true“ oblik. Ovime je završen postupak predprocesuiranja baze podataka i model je spreman za kreiranje frekventnih skupova, a zatim i asocijativnih pravila.

### 5.2.2. Izrada modela za kreiranje asocijativnih pravila

Nakon što je baza podataka uspješno predprocesuirana, pristupa se izradi modela za kreiranje asocijativnih pravila. U našoj bazi podataka potrebna su nam 4 operatora kako bi kreirali model. To su sljedeći operatori:

- Read Excel
- Numerical to Binominal
- FP- Growth
- Create association rules

Na slici 19 je prikazan kreirani model za generiranje asocijativnih pravila.



**Slika 19. Kreirani model za generiranje asocijativnih pravila**

Svaki operator ima svoje značajke, a koje su opisane u sljedećem poglavlju.

#### *Read Excel operator*

Ovaj operator se koristi za učitavanje podataka iz Excel tablice. Korisnik definira koja će se od tablica (sheet) iz cijelog dokumenta koristiti kao podatkovna tablica [11]. Tablica mora biti formatirana na način da svaki njen redak predstavlja primjer, a svaki njen stupac predstavlja alternativu. Treba obratiti pažnju da prvi redak Excel tablice može predstavljati imena atributa, koji mogu biti prikazani kao parametri. [11, 12]

### *Numerical to Binominal operator*

Ovaj operator mijenja tip odabranih numeričkih atributa u binominalan oblik (još se naziva i binarni). Ovaj operator ne samo da mijenja tip odabranih atributa nego bilježi sve vrijednosti tih atributa u odgovarajuće binominalne (binarne) vrijednosti. Binominalni atributi mogu imati samo dvije moguće vrijednosti (npr. „true“ ili „false“, ili 0 ili 1). Ako je vrijednost atributa između postavljene minimalne i maksimalne vrijednosti, ona postaje „true“, ako je suprotno onda je „false“. Minimalne ili maksimalne vrijednosti se postavljaju min i max parametrima. Predefinirane granice za min i max parametre su 0, tako da je 0.0 pretvoreno u „false“, a sve ostale vrijednosti su „true“. [11, 12]

### *FP- Growth operator*

Ovaj operator efikasno izračunava sve frekventne podskupove podataka iz odabranog skupa, koristeći strukturu FP- drva. Nužno je da svi atributi ulaznog skupa podataka budu binominlanog oblika. [11, 12]

### *Create association rules operator*

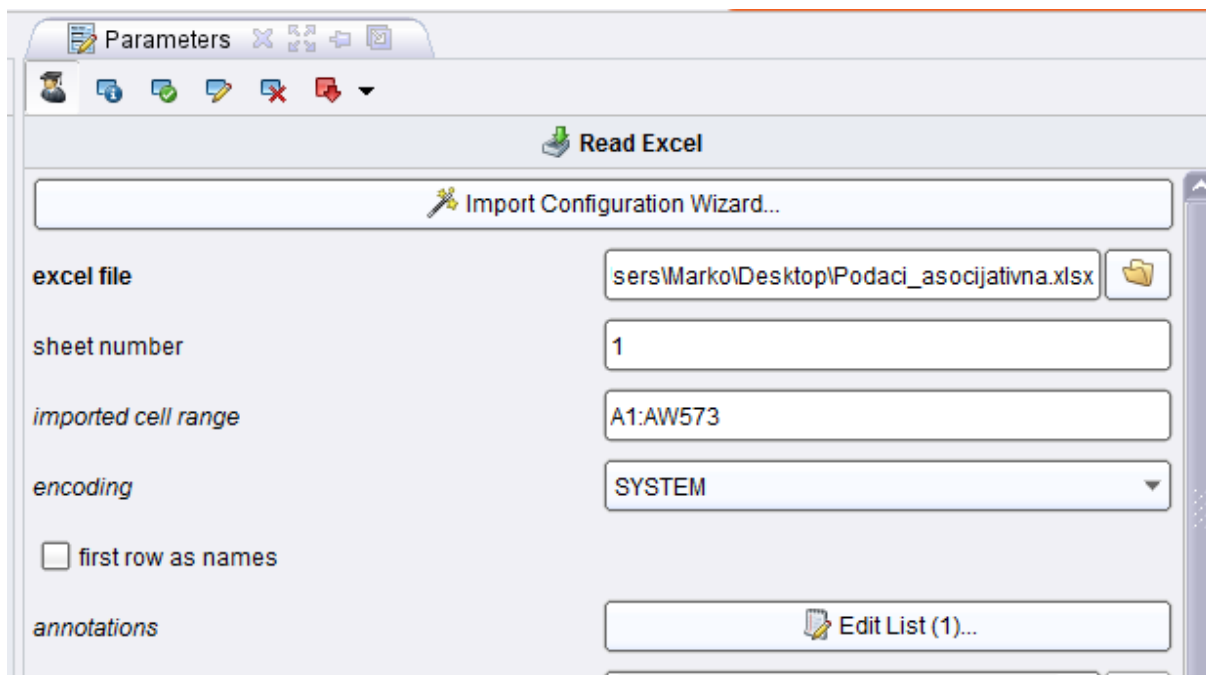
Ovaj operator generira set asocijativnih pravila za odabrani skup frekventnih podskupova podataka. Asocijativna pravila su skup if/then pravila koja pomažu da se otkriju poveznice između prividno nepovezanih podataka. Primjer asocijativnog pravila bi bio: ako kupac kupi jaja, on će sa 80% vjerojatnosti kupiti i mlijeko.

Asocijativno pravilo ima dva dijela: uzrok (eng. *if*), te posljedicu (eng. *then*). Uzrok je stavka iz originalnog (učitanog) skupa podataka. Posljedica je također stavka koja je povezana sa uzrokom.

Asocijativna pravila se kreiraju analizom podataka za česte if/ then uzorke, te koristeći pravila (kriterije), podršku (eng. *Support*) i povjerenje (eng. *Confidence*) identificiraju najznačajnije odnose između njih. Podrška je indikacija koliko često se određena stavka pojavljuje u bazi podataka. Povjerenje je pokazatelj koliko se puta pokazalo da su uzročno/posljedični odnosi točni. Ti odnosi se dobivaju pomoću operatora poput FP- rasta, te operatora Create Association Rules koji uzima generirane frekventne podskupove i pomoću njih kreira asocijativna pravila. [11, 12]

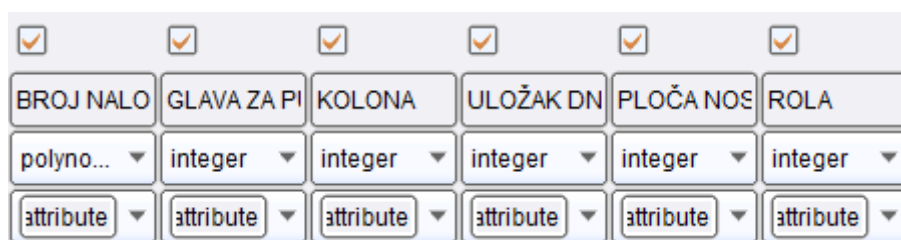
### 5.2.3. Analiza modela i parametara operatora

Nakon izrade modela slijedi učitavanje predprocesuirane baze podataka, optimiranje parametara unutar pojedinih operatora i analiza dobivenih rezultata. Prvo je potrebno učitati bazu podataka pomoću *Read Excel* operatora. Kako bi učitali bazu podataka potrebno je kliknuti na operator te zatim pokrenuti čarobnjak za učitavanja baze podataka kako je prikazano na slici 20.



Slika 20. Podešavanje *Read Excel* operatora

Pokretanjem čarobnjaka otvara se novi prozor u kojemu biramo .xlsx datoteku koju želimo učitati. Nakon toga odabiremo koju radnu stranicu datoteke trebamo analizirati, koje redove želimo označiti kao imena atributa te na kraju tip pojedinih vrijednosti. Na slici je 21 je prikazan tip vrijednosti koje predlaže programski paket RapidMiner.



Slika 21. Predloženi tip vrijednosti

RapidMiner predlaže tip atributa „*integer*“, što ne odgovara našem modelu. Taj problem rješavamo pomoću prije spomenutog operatora „*Numerical to Binominal*“, koji će sve cjelobrojne vrijednosti u bazi podataka pretvoriti u binominalni oblik. Na slikama 22 i 23 prikazane su razlike u tipu vrijednosti nakon uključivanja operatora.

BROJ NALOGA	Polynomial
GLAVA ZA PUHANJE	Integer
KOLONA	Integer
ULOŽAK DNA KALUPA	Integer

Slika 22. Tip vrijednosti prije obrade podataka

GLAVA ZA PUHANJE	Binominal
KOLONA	Binominal
ULOŽAK DNA KALUPA	Binominal
PLOČA NOSAČA KALUPA	Binominal

Slika 23. Tip vrijednosti nakon obrade pomoću operatora „*Numerical to Binominal*“

Baza podataka je sada spremna za sljedeći operator. Idući operator je „*FP-Growth*“, koji će generirati frekventne podskupove. Pomoću frekventnih podskupova vidjeti će se koje pozicije se pojavljuju dovoljan broj puta kako bi zadovoljile uvjet potpore. Pojam potpore (eng. *support*) i povjerenja (eng. *confidence*) objašnjen je u sljedećem poglavlju. Za ovaj primjer odabrana je vrijednost potpore od 0.5, kako je prikazano na slici 24. Vrijednost povjerenja definira se pomoću parametra u operatoru za kreiranje asocijativnih pravila.

Parameters

FP-Growth

find min number of itemsets

min number of itemsets: 100

max number of retries: 15

positive value: 1

min support: 0.5

max items: -1

must contain:

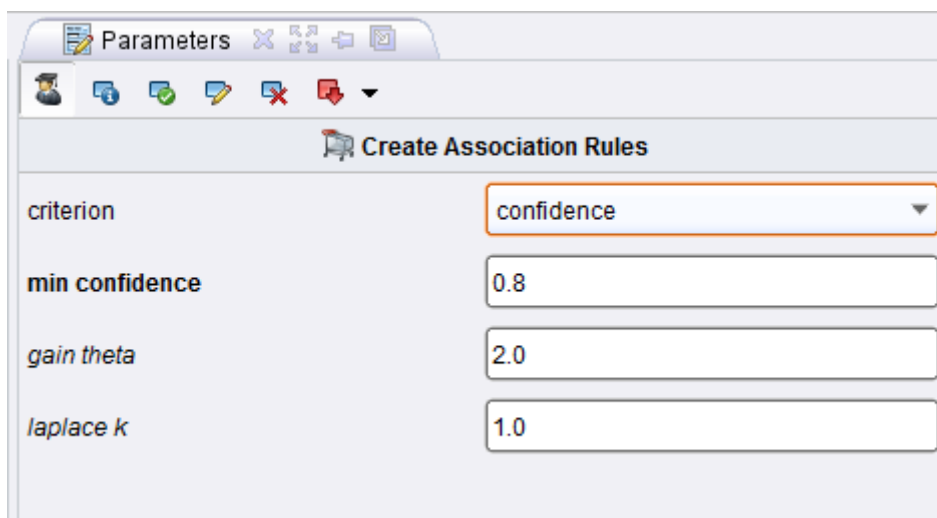
Slika 24. Postavke „FP- Growth“ operatora

Na slici 25 prikazani su generirani frekventni podskupovi, uz definiranu minimalnu potporu u iznosu od 0.5.

Size	Support	Item 1	Item 2	Item 3	Item 4
1	0.731	DNO KALUPA			
1	0.687	KALUP			
2	0.678	DNO KALUPA	KALUP		
1	0.656	PRETKALUP			
1	0.596	DNO PRETKALUPA			
2	0.587	PRETKALUP	DNO PRETKALUPA		
2	0.570	DNO KALUPA	PRETKALUP		
2	0.558	KALUP	PRETKALUP		
3	0.552	DNO KALUPA	KALUP	PRETKALUP	
2	0.531	DNO KALUPA	DNO PRETKALUPA		
3	0.530	DNO KALUPA	PRETKALUP	DNO PRETKALUPA	
2	0.514	KALUP	DNO PRETKALUPA		
3	0.514	DNO KALUPA	KALUP	DNO PRETKALUPA	
3	0.512	KALUP	PRETKALUP	DNO PRETKALUPA	
4	0.512	DNO KALUPA	KALUP	PRETKALUP	DNO PRETKALUPA

Slika 25. Generirani frekventni podskupovi

Nakon operatora za kreiranje frekventnih podskupova slijedi operator za kreiranje asocijativnih pravila. Pomoću ovog operatora kreiraju se asocijativna pravila koja nam govore koje su stavke na naložima povezana. Na slici 26 prikazane su postavke algoritma. Za početnu vrijednost povjerenja odabrana je vrijednost od 0.8.



Slika 26. Postavke operatora za kreiranje asocijativnih pravila

Analiza rezultata slijedi nakon objašnjenja postavka algoritma, kao i kriterija prema kojima razlikujemo asocijativna pravila.

*Povjerenje (eng. Confidence)*

Povjerenje nekog pravila tumači se kao uvjetna vjerojatnost da će neka transakcija sadržavati stavku Y ukoliko znamo da sadrži stavku X. Povjerenje se tako računa prema izrazu:

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (1)$$

Gdje je  $\sigma(X \cup Y)$  ukupan broj transakcija koje sadrže stavke X i Y, a  $\sigma(X)$  ukupan broj transakcija koje sadrže stavku X. Povjerenje se može shvatiti i kao mjera snage asocijacije u pravilu. [13]



### Potpora (eng. *Support*)

Potpora se definira kao vjerojatnost da će slučajno odabrana transakcija sadržavati stavke X i Y. Potpora sadrži posljedicu i pretpostavku, te također eliminira pravila koja se rijetko ponavljaju. Potpora se računa prema sljedećem izrazu:

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (2)$$

gdje je  $\sigma(X \cup Y)$  ukupan broj transakcija koje sadrže stavke X i Y i naziva se još i potporni broj (eng. *Support count*), dok N označava ukupan broj transakcija u bazi podataka. [13]

### Conviction pravila

Conviction pravila se definira prema izrazu:

$$conv(X \rightarrow Y) = \frac{1 - supp(Y)}{1 - conf(X \rightarrow Y)} \quad (3)$$

i tumači se kao odnos očekivane frekvencije da se stavka X javlja bez stavke Y (tj. da će pravilo biti pogrešno) i frekvencije da se X javlja bez Y dobivene na osnovu podataka. Ako bi ova vrijednost iznosila 1,5, to pravilo bilo bi 1,5 puta češće pogrešno da su X i Y stavke nezavisne, što znači da je pravilo interesantno. [13]

### Lift pravila

Lift pravila tumači se kao odnos potpore dobivene na osnovi podataka i očekivane potpore ako X i Y ne ovise jedan o drugom, a računa se prema izrazu:

$$lift(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(Y) \times supp(X)} \quad (4)$$

U praktičnim primjerima značajnost *lift-a* se tumači kao pozitivna veza ako je vrijednost *lift-a* veća od 1, odnosno da je vjerojatno da će se na nalogu naći stavka Y ako se nalogu nalazi stavka X. Negativna veza vrijedi za vrijednost *lift-a* manju od 1, a znači da je vjerojatno da se na nalogu neće naći stavka Y ako na nalogu postoji stavka X. Ukoliko je vrijednost *lift-a* jednaka 1, to znači da su dvije stavke nezavisne, odnosno da je pravilo beznačajno. [13]

#### 5.2.4. Analiza dobivenih asocijativnih pravila

Dobivena asocijativna pravila smo dobili tako da smo izlaz operatora za kreiranje asocijativnih pravila spojili na izlaz procesa. Tražimo pravila koja imaju potporu veću od 0.5 i povjerenje veće od 0.8. Na slici 27 su prikazana dobivena asocijativna pravila.

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
189	DNO KALUPA	KALUP	0.678	0.928	0.970	-0.783	0.176	1.351	4.360
626	KALUP	DNO KALUPA	0.678	0.987	0.995	-0.696	0.176	1.351	21.162
537	DNO PRETKALUPA	PRETKALUP	0.587	0.985	0.995	-0.605	0.197	1.503	23.488
446	DNO KALUPA, PRETKALUP	KALUP	0.552	0.969	0.989	-0.587	0.161	1.411	10.202
664	KALUP, PRETKALUP	DNO KALUPA	0.552	0.991	0.997	-0.563	0.145	1.356	28.628
6	PRETKALUP, DNO PRETKALUPA	DNO KALUPA	0.530	0.902	0.964	-0.645	0.100	1.234	2.741
199	DNO KALUPA, PRETKALUP	DNO PRETKALUPA	0.530	0.929	0.974	-0.610	0.190	1.559	5.724
777	DNO KALUPA, DNO PRETKALUPA	PRETKALUP	0.530	0.997	0.999	-0.533	0.181	1.520	104.699
433	DNO KALUPA, DNO PRETKALUPA	KALUP	0.514	0.967	0.989	-0.549	0.149	1.408	9.513
778	KALUP, DNO PRETKALUPA	DNO KALUPA	0.514	1	1	-0.514	0.138	1.368	$\infty$
107	KALUP, PRETKALUP	DNO PRETKALUPA	0.512	0.918	0.971	-0.603	0.180	1.541	4.955
108	KALUP, PRETKALUP	DNO KALUPA, DNO PRETKALUPA	0.512	0.918	0.971	-0.603	0.216	1.728	5.749
176	DNO KALUPA, KALUP, PRETKALUP	DNO PRETKALUPA	0.512	0.927	0.974	-0.593	0.183	1.555	5.548
394	DNO KALUPA, DNO PRETKALUPA	KALUP, PRETKALUP	0.512	0.964	0.987	-0.551	0.216	1.728	12.224
432	DNO KALUPA, PRETKALUP, DNO PRETKALUPA	KALUP	0.512	0.967	0.989	-0.547	0.148	1.407	9.482
774	KALUP, DNO PRETKALUPA	PRETKALUP	0.512	0.997	0.999	-0.516	0.175	1.520	101.255
775	KALUP, DNO PRETKALUPA	DNO KALUPA, PRETKALUP	0.512	0.997	0.999	-0.516	0.219	1.749	126.441
776	DNO KALUPA, KALUP, DNO PRETKALUPA	PRETKALUP	0.512	0.997	0.999	-0.516	0.175	1.520	101.255
782	KALUP, PRETKALUP, DNO PRETKALUPA	DNO KALUPA	0.512	1	1	-0.512	0.138	1.368	$\infty$

Slika 27. Dobivena asocijativna pravila

Slijedi analiza dobivenih asocijativnih pravila. Sa slike vidimo kako su dobiveni skupovi koji imaju i traženu potporu i povjerenje. Ako neki skup ima potporu veću od 0,5, to znači da se taj skup pojavio u prvotnoj bazi podataka barem u 50% zapisa. Konkretno za ovaj slučaj, to znači da bi naši skupovi prošli ovaj uvjet moraju se pojaviti najmanje 286 puta u originalnoj bazi podataka. Povjerenje se ovdje tumači kao pouzdanost da je premisa točna, tj. da je asocijativno pravilo ispravno. Tako se, na primjer, za prvo asocijativno pravilo sa slike 4-18 može tvrditi da je ispravno u 92,8% slučajeva. To znači da se u 92,8% zapisa koji su sadržavali poziciju DNO KALUPA nalazila i pozicija KALUP. Na isti način se tumače i ostala dobivena asocijativna pravila. Dvije kategorije koje su također zanimljive su ranije spomenuti „lift“ i „conviction“. Za sva pravila možemo primijetiti kako je vrijednost lift-a veća od 1, što upućuje na pozitivnu vezu. To znači da je, na primjer, za prvo pravilo, vjerojatno da će se na nalogu pojaviti pozicija DNO KALUPA ako se na nalogu već nalazi pozicija KALUP. Također su interesantna i pravila koja imaju vrijednost potpore veću od traženih 0, 5 i vrijednost povjerenja jednaku 1. Ovakva pravila su jednoznačna, tj. kod njih se

uvijek pojavljuje pozicija koja je uvjetovana nekom premisom. Za pravilo koje kaže da se DNO KALUPA i DNO PRETKALUPA pojavljuje na naložima na kojima se nalaze KALUP i PRETKALUP možemo reći kako je u potpunosti ispravno. Iako su se ovi zapisi pojavljivali u originalnoj bazi podataka više od 286 puta, u svakom slučaju je ovo pravilo bilo ispravno. Ovo je značajan podatak za proizvodno poduzeće zbog toga što poduzeće sada zna koji proizvodi uvjetuju neke druge proizvode. Kod ovog pravila također vidimo da je vrijednost conviction-a beskonačna, što znači da bi pravilo bilo pogrešno u slučaju da ove pozicije nisu povezane. Budući da znamo kako je pravilo ispravno, zaključak je da su ove 4 pozicije usko povezane i ovise jedna od drugoj. Za ostala pravila također se može reći isto, jer vrijednost conviction-a je za svako pravilo značajno veća od 1.

**AssociationRules**

```

Association Rules
[PRETKALUP, DNO PRETKALUPA, CANGA] --> [JEZGRENİK] (confidence: 0.901)
[PRETKALUP, DNO PRETKALUPA, CANGA] --> [PRSTEN, JEZGRENİK] (confidence: 0.901)
[DNO KALUPA, PRSTEN, CANGA] --> [KALUP, DNO PRETKALUPA] (confidence: 0.901)
[DNO KALUPA, PRSTEN, CANGA] --> [KALUP, PRETKALUP, DNO PRETKALUPA] (confidence: 0.901)
[PRETKALUP, DNO PRETKALUPA, PRSTEN] --> [JEZGRENİK] (confidence: 0.901)
[PRETKALUP, DNO PRETKALUPA] --> [DNO KALUPA] (confidence: 0.902)

```

**Slika 28. Asocijativna pravila poredana prema povjerenju**

**Tablica 6. Asocijativna pravila i vrijednosti značajnosti pravila**

Uvjet	Zaključak	Potpورا	Povjerenje	Lift	Conviction
DNO KALUPA	KALUP	0,678	0,928	1,351	4,36
KALUP	DNO KALUPA	0,678	0,987	1,351	21,162
DNO PRETKALUPA	PRETKALUP	0,587	0,985	1,503	23,488
DNO KALUPA, PRETKALUP	KALUP	0,552	0,969	1,411	10,202
KALUP, PRETKALUP	DNO PRETKALUPA	0,552	0,991	1,356	28,62

DNO KALUPA, PRETKALUP	PRETKALUP	0,530	0,929	1,559	5,724
DNO KALUPA, DNO PRETKALUPA	KALUP	0,530	0,997	1,520	104,6
DNO KALUPA, DNO PRETKALUPA	DNO KALUPA	0,514	0,967	1,408	9,513
KALUP, DNO PRETKALUPA	DNO PRETKALUPA	0,514	1	1,368	∞
DNO KALUPA, KALUP, PRETKALUP	KALUP, PRETKALUP	0,512	0,927	1,555	5,548
DNO KALUPA, DNO PRETKALUPA	KALUP	0,512	0,964	1,728	12,22
DNO KALUPA, PRETKALUP, DNO PRETKALUPA	PRETKALUP	0,512	0,967	1,407	9,482
KALUP, DNO PRETKALUPA	PRETKALUP	0,512	0,997	1,520	101,2
DNO KALUPA, KALUP, DNO PRETKALUPA	DNO KALUPA	0,512	0,997	1,749	126,4
KALUP, PRETKALUP, DNO PRETKALUPA	DNO PRETKALUPA	0,512	0,997	1,520	101,2

### 5.3. Modeliranje poslovnih pravila

Poslovno pravilo je pravilo poduzeća ili korporacije. To je pravilo koje definira ili ograničava neke aspekte poslovanja. Poslovna pravila su namijenjena da nametnu poslovnu strukturu ili da upravljaju ili utječu na ponašanje poduzeća. Poslovna pravila opisuju operacije, definicije i ograničenja koja se primjenjuju na organizaciju. Mogu se primijeniti na

ljude, procese i računalne sustave u organizaciji i donose se kako bi pomogle organizaciji da postigne svoje ciljeve. [6, 7]

Poslovno pravilo predstavlja onaj dio znanja organizacije koji je zapisan, tj. koji je eksplicitno prikazan. Implicitno znanje, koje zaposlenici imaju kao dio svog iskustva i osobnog znanja, a koje nije nigdje fizički izraženo, neovisno o utjecaju na poslovanje, ne predstavlja poslovna pravila zato što nije dostupno svima. Takvo znanje potrebno je izraziti eksplicitno, kako bi ono postalo dio poslovnih pravila organizacije. Razlog tome je činjenica da implicitno znanje zaposlenici „nose“ sa sobom, čime se iznimno ograničava iskoristivost tog znanja, kao i rizik od nestajanja istog. Primjerice, ekspert iz područja javne nabave, u nekoj organizaciji, koji znanje drži za sebe, bez njegove eksplikacije, može svoju organizaciju napustiti, čime povlači i sve svoje znanje potrebno za kvalitetan rad organizacije. Da je takvo znanje eksplicitno izraženo, npr. zapisano u obliku nekih pravila, iskustva ili uputa, znanje bi lakše ostalo u organizaciji čak i kad je ekspert napusti. Poslovna pravila bi trebala biti minimalna, tj. da ih nije moguće razbiti na više manjih pravila. [6, 7]

Osnovna načela poslovnih pravila:

- Niti jedno poslovno pravilo nije vječno. Poslovna pravila podložna su izmjenama. Činjenica je da se poslovanje organizacije mijenja kroz vrijeme, pa je potrebno sukladno tome mijenjati i poslovna pravila
- Poslovno pravilo mora imati globalnog smisla, kroz cijeli arhitekturni okvir
- Nema poslovnih pravila sve dok se ne kaže da ona postoje
- Poslovno pravilo znači upravo ono što je određeno korištenim riječima – ništa više, ništa manje
- Poslovno pravilo je uvijek primjenjivo. Poslovno pravilo tako je postavljeno da je uvijek moguće točno odrediti njegovu primjenjivost, odnosno uvijek je jasno na što se poslovno pravilo odnosi. [6, 7]

Poslovna pravila biti će modelirana pomoću dva operatora. Riječ je o prediktivnim modelima koji služe za procjenu nekog parametra. U bazi podataka postoji polje faktor cijene koje je zanimljivo jer će se pomoću operatora u programskom paketu RapidMiner htijeti pronaći odgovarajući model kojime će se moći procijeniti cijena nekog proizvoda ili pozicije na temelju postojećih podataka. Drugim operatorom želimo predvidjeti državu isporuke. U ovom slučaju ne možemo koristiti neuronske mreže jer varijabla koju želimo predvidjeti nije

numerička, već tekstualna. U tom slučaju nam je potreban “*Rule Induction*” operator koji će pomoću niza testiranih pravila i diskretiziranih vrijednosti podataka koji ulaze u analizu kreirati poslovna pravila pomoću kojih procjenjujemo državu isporuke.

### **5.3.1. Izrada prediktivnog modela pomoću neuronskih mreža**

Prediktivni modeli su izrađeni pomoću dva operatora. U radu je prvo opisan model kojime želimo predvidjeti faktore cijene. Neuronske mreže u programskom sustavu RapidMiner rade samo s numeričkim vrijednostima pa će svi podaci koji ulaze u razmatranje morati biti numeričkog oblika. U sljedećim poglavljima su opisani operatori koji su nam potrebni za kreiranje modela, a nismo ih koristili kod modela za kreiranje asocijativnih pravila.

#### *Operator za odabir atributa*

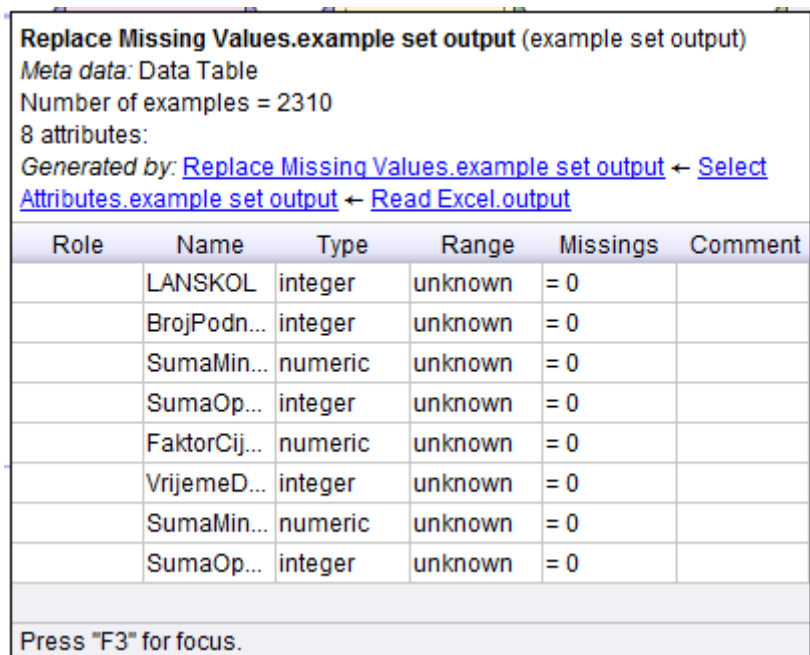
Operator za odabir atributa (eng. *Select Attributes*) služi nam kako bi odabrali attribute koje želimo uzeti u razmatranje za analizu. U našem primjeru tako pomoću operatora za učitavanje baze podataka učitavamo sve podatke. Nakon toga koristimo operator za odabir atributa i pomoću njega odabiremo samo attribute koji će nam služiti za kreiranje modela. Ovdje su to sljedeći atributi:

- Broj podnaloga
- Faktor cijene
- Lansirana količina
- Suma operacija na podnalogu
- Suma operacija na nalogu
- Suma minuta na podnalogu
- Suma minuta na nalogu
- Vrijeme do isporuke

U teoriji, ako u analizi uzimamo više atributa, trebali bi dobiti točniji model. Zbog toga su u slučaju operatora za kreiranje modela pomoću neuronskih mreža korišteni svi numerički atributi iz baze podataka. [11, 12]

### Operator za popunjavanje praznih vrijednosti

Ovaj operator nam je potreban kako bi zamijenili prazne vrijednosti iz baze podataka nekim drugim vrijednostima. Operator za kreiranje neuronskih mreža ne može raditi ukoliko su umjesto nekih vrijednosti u bazi podataka samo prazna polja. Sa slike 29 vidimo izlaz sa operatora za popunjavanje praznih vrijednosti (eng. *Replace missing values*). Sada je baza spremna za daljnje predprocesuiranje. [11, 12]



Role	Name	Type	Range	Missings	Comment
	LANSKOL	integer	unknown	= 0	
	BrojPodn...	integer	unknown	= 0	
	SumaMin...	numeric	unknown	= 0	
	SumaOp...	integer	unknown	= 0	
	FaktorCij...	numeric	unknown	= 0	
	VrijemeD...	integer	unknown	= 0	
	SumaMin...	numeric	unknown	= 0	
	SumaOp...	integer	unknown	= 0	

Slika 29. Izlaz operatora za popunjavanje praznih vrijednosti

### Operator za postavljanje uloge atributa

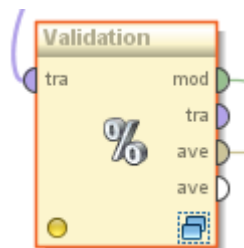
Operator za postavljanje uloge atributa (eng. *Set role*) služi nam kako bi stavili oznaku na atribut koji želimo iskoristiti u trening i test setu za predviđanje. Sa neuronskim mrežama želimo znati koliki će biti faktor cijene na temelju kreiranog modela. Zbog toga je potrebno staviti oznaku na polje faktor cijene kako je prikazano na slici 30. [11, 12]



Slika 30. Postavljanje oznake na faktoru cijene

### *Operator za izvršavanje unakrsne validacije*

Ovaj operator izvršava unakrsnu validaciju modela. Riječ je o operatoru koji unutar sebe ima dva podprocesa. To su trening i test procesi. Unutar trening procesa RapidMiner kreira model koji se zatim testira. Također se mjeri i analizira dobrotu prilagodbe trening modela test modelu. Nakon toga se izvršava unakrsna validacija kojom se želi provjeriti koliko će dobro kreirani model raditi u praksi. [11, 12]



**Slika 31. Operator za unakrsnu validaciju modela**

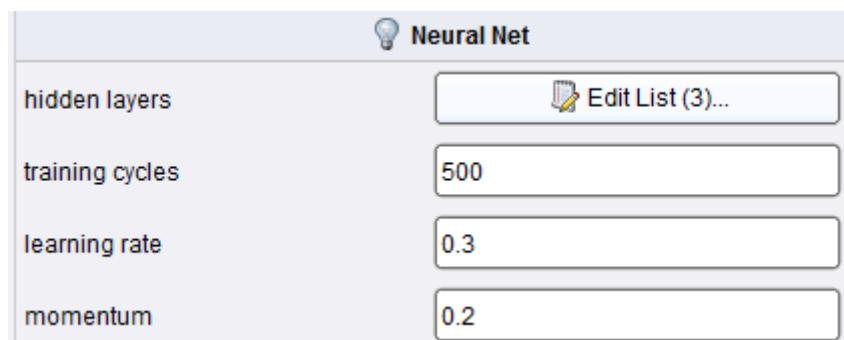
U parametrima je moguće odabrati željeni broj validacija. U teoriji, veći broj validacija bi trebao značiti i bolji model, ali veći broj validacija značajno usporava cijeli proces. Za bazu podataka u kojoj želimo predvidjeti faktor cijene pozicije odabrano je 10 validacija.

### *Operator za treniranje neuronske mreže*

Operator nam služi za treniranje neuronske mreže. Radi na principu unaprijedne neuronske mreže s povratnim prostiranjem pogreške i ima mogućnost učenja. U parametrima neuronske mreže je moguće podesiti broj skrivenih slojeva mreže, momentum i koeficijent učenja. Odabrane su vrijednosti prikazane na slici 32. Također je bitno napomenuti kako je uključena opcija normalizacije podataka što povećava opterećenje računala, ali je potrebno kako bi se podaci postavili na normalizirane vrijednosti (između 0 i 1). [11, 12]. Odabrano je 500 ciklusa treninga neuronske, koeficijent učenja 0,3 i moment učenja 0,2. Ove vrijednosti su odabrane kako bi se kasnije upotrebom operatora za unakrsnu validaciju osigurala maksimalna točnost testirane neuronske mreže. Unakrsnom validacijom osigurava se prekid treniranja neuronske mreže ukoliko vrijednost pogreške dosegne odabranu razinu. Korištene je unaprijedna neuronska mreže koja ima mogućnost učenja. Kod unaprijedne neuronske mreže postoji samo jedan smjer kretanja težina neurona, i to od početnog neurona, preko skrivenih slojeva pa do izlaznog neurona. Koeficijent učenja je vrijednost za koju se mijenja težina pojedinih neurona u svakom koraku učenja, a moment dodaje dio vrijednosti promjene



težine neurona iz prethodnog koraka težini neurona u sljedećem koraku. Ovime se neuronska mreža osigurava od prekida u slučaju pronalaženja lokalnog minimuma ili maksimuma.



**Slika 32. Postavke Neural Net operatora**

#### *Operator za mjerenje performansi modela*

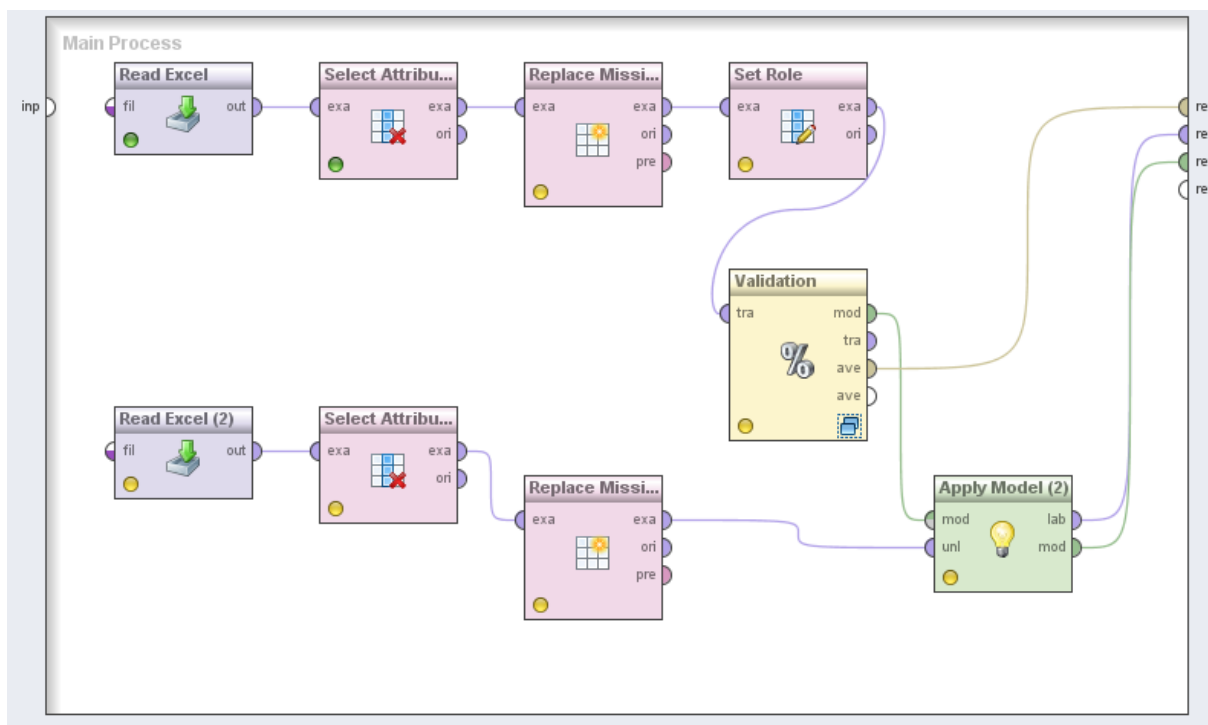
Ovaj operator nam služi kako bi dobili vektor performansi koji nam govori koliko je dobar naš model u praksi. Pomoću vektora performansi (eng. *Performance Vector*) moguće je očitati niz statističkih podataka koji su posljedica kreiranog modela. Pomoću ovih podataka možemo pretpostaviti koliko će dobro naš model predviđati podatke koji nisu u test i trening setu. [11, 12]

#### *Operator za primjenu modela*

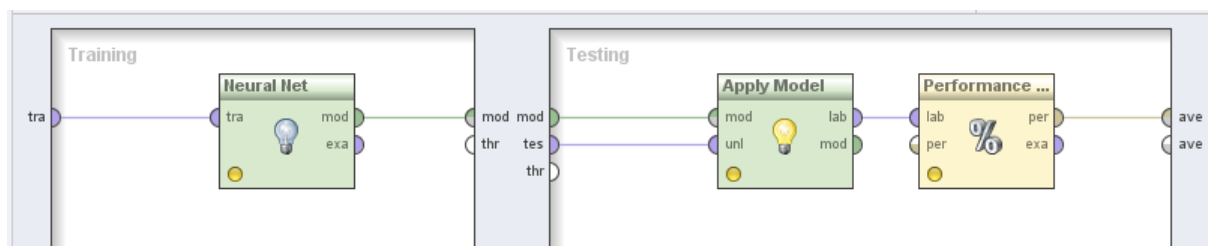
Operator za primjenu modela (eng. *Apply Model*) služi kako bi istrenirani ili testirani model iskoristili na novom setu podataka nakon čega možemo vidjeti koliko dobro model stvarno radi. [11, 12]

### **5.3.2. Analiza kreiranog modela i dobivenih rezultata**

Na slici 33 je prikazan kreirani prediktivni model pomoću neuronskih mreža. Operator za kreiranje neuronske mreže je unutar operatora za izvršavanje unakrsne validacije. Neuronska mreža je unutar prvog podprocesa i služi za treniranje modela. U drugom podprocesu vidimo operatore za mjerenja performansi modela i operator za primjenu istreniranog modela.



Slika 33. Kreirani glavni proces



Slika 34. Trening i test podproces

Nakon kreiranja modela krećemo u analizu rezultata. Kreirani model prvo predprocesuira bazu podataka, zatim odabire polje koje želimo predvidjeti, te na kraju primjenjuje istrenirani model na novom setu podataka. Na kraju su spojeni svi bitni izlazi operatora iz kojih možemo iščitati korisne podatke o modelu. Na slici 35 prikazan je vektor performansi iz kojega očitavamo neke bitne karakteristike pomoću kojih možemo procijeniti koliko je model dobar.

## PerformanceVector

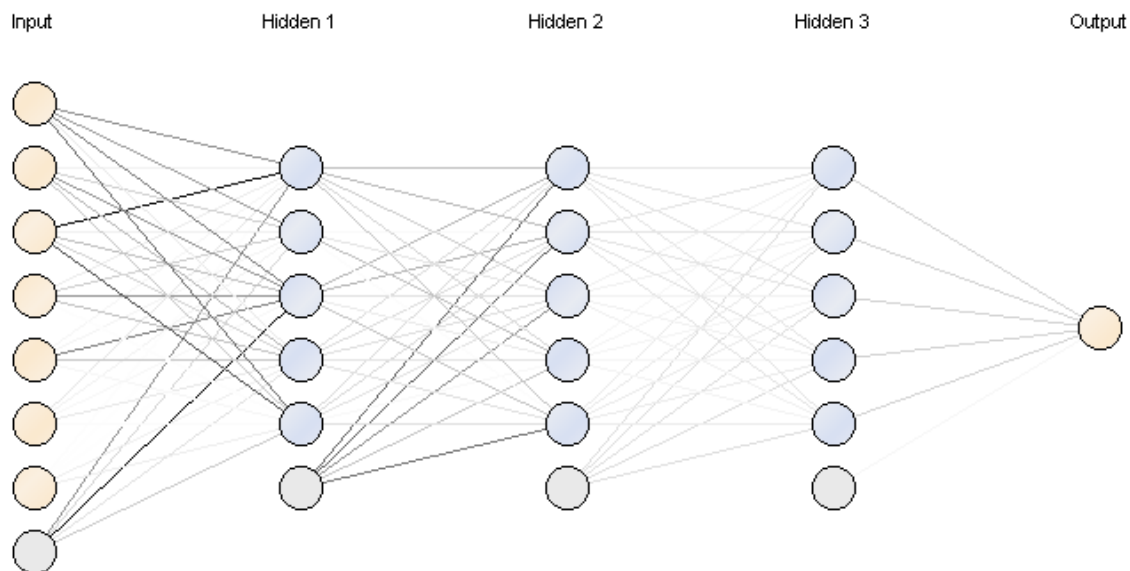
```
PerformanceVector:  
root_mean_squared_error: 0.956 +/- 0.185 (mikro: 0.974 +/- 0.000)  
normalized_absolute_error: 0.440 +/- 0.150 (mikro: 0.438)  
root_relative_squared_error: 0.565 +/- 0.072 (mikro: 0.572)  
squared_error: 0.949 +/- 0.375 (mikro: 0.949 +/- 4.779)  
correlation: 0.853 +/- 0.018 (mikro: 0.829)  
squared_correlation: 0.727 +/- 0.031 (mikro: 0.688)
```

**Slika 35. Vektor performansi**

Sa vektora performansi vidimo kako bi model u većini slučajeva dobro predviđao vrijednost faktora cijene. Najvažnija komponenta vektora performansi je koeficijent determinacije koji ovdje iznosi 0,727, što je vrlo dobar rezultat. To znači da je modelom objašnjeno 72,7% varijacije ukupnog skupa. Koeficijent korelacije je također vrlo dobar, što ukazuje na povezanost varijabli uključenih u analizu sa faktorom cijene. Na slici 36 prikazan je dio predviđenih vrijednosti novog skupa podataka uz pomoć kreiranog modela. Također je na slici 37 prikazan i model neuronskih mreža koji nam je poslužio za treniranje modela. Vidi se kako neuronska mreža ima tri skrivena sloja, 7 ulaznih parametara i jedan izlazni.

Row No.	prediction(F...	LANSKOL	BrojPodnal...	SumaMinut...	SumaOper...	FaktorCijene	vrijemeDol...	SumaMinut...	SumaOper...
465	2.794	51	6	53598.800	125	2.196	5	2645	13
466	2.580	24	6	43021.200	146	2.153	8	4980	19
467	2.907	1	6	43021.200	146	2.153	8	579	23
468	2.685	24	6	43021.200	146	2.153	8	11046	26
469	2.443	24	6	43021.200	146	2.153	8	1354	11
470	2.871	24	6	43021.200	146	2.153	8	21478.200	42
471	2.673	24	6	43021.200	146	2.153	8	3584	25
472	0.353	17	6	11919.380	136	0.653	9	967	14
473	0.385	12	6	11919.380	136	0.653	9	1180	12
474	0.500	3	6	11919.380	136	0.653	9	3814.200	36
475	0.391	13	6	11919.380	136	0.653	9	2340.060	26
476	0.443	3	6	11919.380	136	0.653	9	933.120	18
477	0.475	3	6	11919.380	136	0.653	9	2685	30
478	0.274	1	10	6847.800	164	0.260	4	600	19
479	0.272	3	10	6847.800	164	0.260	4	1012	16
480	0.272	2	10	6847.800	164	0.260	4	400	14
481	0.277	1	10	6847.800	164	0.260	4	1729.800	31
482	0.276	1	10	6847.800	164	0.260	4	1368	26

Slika 36. Predviđene vrijednosti uz pomoć modela



Slika 37. Struktura primijenjene neuronske mreže za treniranje modela za predviđanje faktora cijene

Na slici 38 prikazane su vrijednosti težina izlaznih neurona. Težina pojedinih neurona predstavlja jačinu veze između ta dva neurona. Tako pomoću težina izlaznog neurona

možemo vidjeti koliki utjecaj imaju pojedini neuroni u zadnjem skrivenom sloju neuronske mreže na ukupni rezultat. Vidi se kako svi parametri vrlo slično utječu na izlazni neuron.

```
Output
=====

Regression (Linear)
-----
Node 1: -2.718
Node 2: -2.725
Node 3: -2.728
Node 4: -2.720
Node 5: -2.726
Threshold: 0.514
```

Slika 38. Težine izlaznih neurona

### 5.3.3. Modeliranje poslovnih pravila pomoću induktivnog učenja

Modeliranjem poslovnih pravila pomoću induktivnog učenja možemo predvidjeti i tekstualne atribute za razliku od modeliranja pomoću neuronskih mreža gdje možemo predvidjeti samo numeričke atribute. Ovu mogućnost želimo iskoristiti kako bi predvidjeli državu isporuke na temelju diskretiziranih numeričkih podataka.

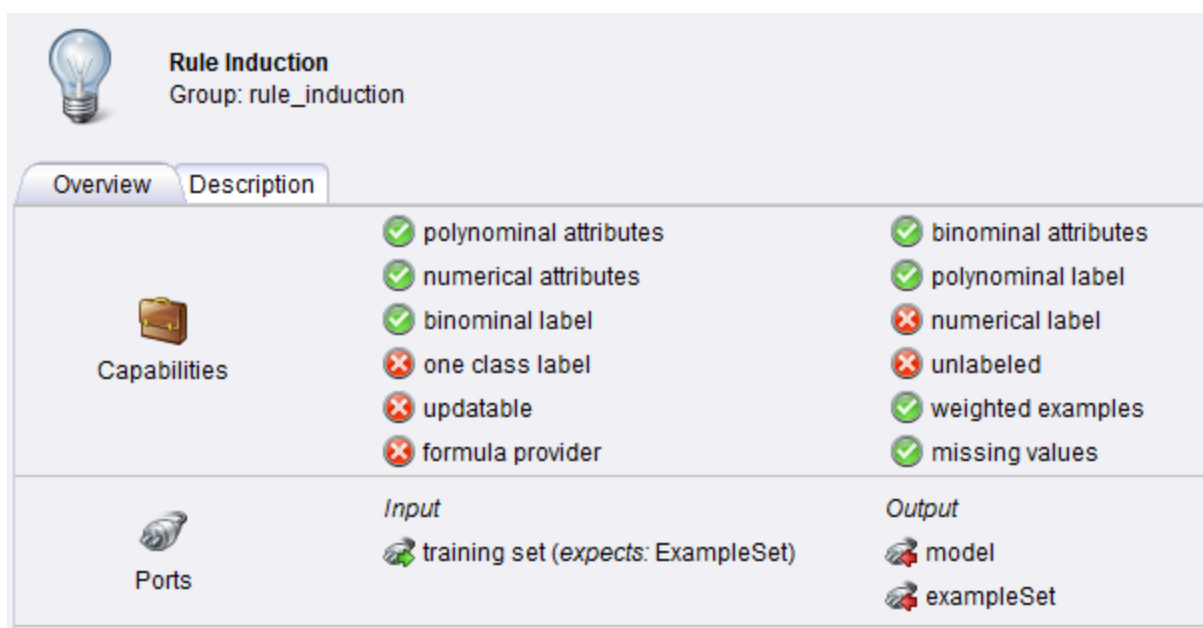
Kako bi kreirali model za induktivno učenje potrebna su nam još dva operatora uz sve već prije navedene prilikom kreiranja modela za učenje pomoću neuronskih mreža. To su operator za diskretizaciju vrijednosti (eng. *Discretize*) i operator za kreiranje poslovnih pravila (eng. *Induction Rule*). Pomoću ova operatora diskretiziramo vrijednosti temeljem broja podataka u bazi. Formula prema kojoj tražimo broj razreda glasi:

$$1 + 3,3 * \log(N) \quad (5)$$

gdje je N ukupan broj podataka. U ovom slučaju N je 2500 zbog toga što je toliko podataka odabrano za treniranje modela. Uvrštavanjem u gornji izraz dobivamo 12,21 razreda, što znači da nam je potrebno 13 razreda podataka. Slijedi analiza novih operatora, modela i dobivenih rezultata.

### Operator za kreiranje poslovnih pravila

Induction rule operator kreira model koji zatim testiramo na isti način kao i model s neuronskim mrežama. Razlika je u tipu podataka s kojima rade ovi operatori. Induction rule operatorom želimo predvidjeti državu isporuke na temelju istih podataka koje smo koristili s neuronskim mrežama. Rule induction operator radi s numeričkim i polinomialnim atributima, a također može raditi i s polinomialnim oznakama, što znači da je moguće predvidjeti tekstualne attribute, kao što je država isporuke. Na slici su prikazane navedene mogućnosti operatora. [11, 12]

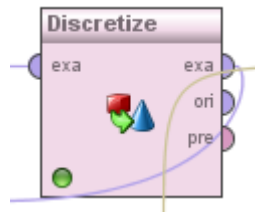


**Slika 39. Mogućnosti Rule Induction operatora**

Ovaj operator radi na temelju modificiranog RIPPER algoritma. Algoritam kreće s manje relevantnim klasama te zatim iterativno raste i obrezuje dobivena pravila sve dok ne ukloni pozitivne primjere ili greška algoritma ne bude veća od 50%.

### Operator za provođenje diskretizacije

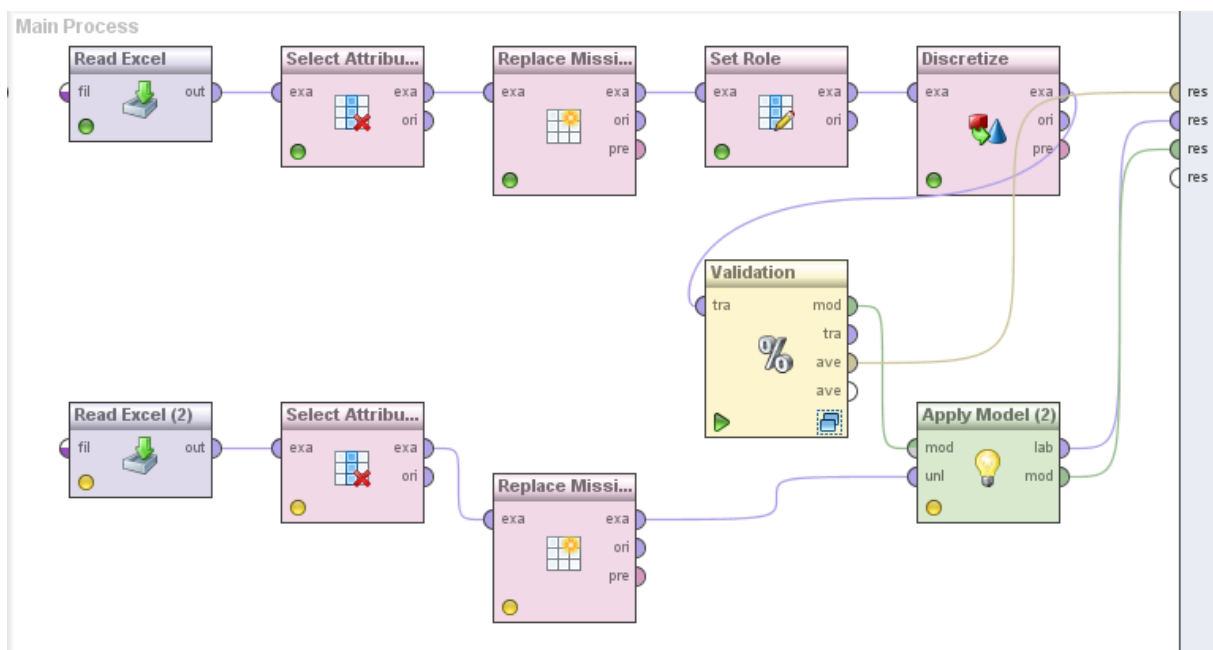
Ovim operatorom želimo provesti jednostavnu podjelu podataka u traženi broj razreda. Postoji više mogućnosti diskretizacije podataka, a nama je bila potrebna samo jednostavna diskretizacija prema veličini podataka. Broj razreda je odabran prema prije navedenoj formuli na temelju broja podataka koji ulaze u trening set. [11, 12]



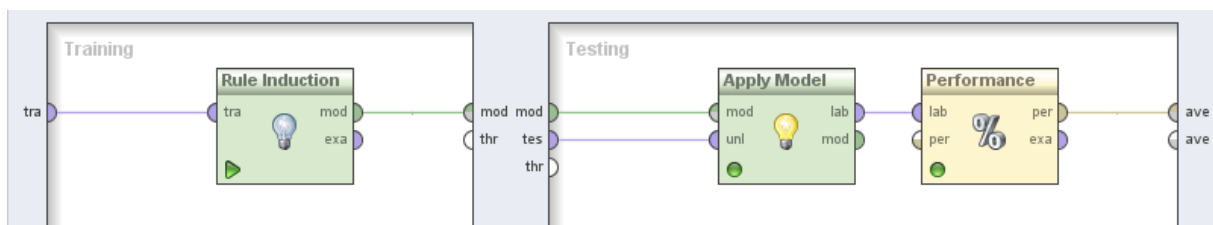
Slika 40. Operator za provođenje diskretizacije

*Analiza modela i rezultata*

Na slikama 41 i 42 su prikazani model i dva podprocesa u modelu.



Slika 41. Glavni proces



Slika 42. Trening i test podproces

Model započinje učitavanjem baze podataka na istom principu kao i u prijašnjim primjerima, zatim biramo atribute koji ulaze u analizu te popunjavamo prazne vrijednosti. Zatim biramo atribut koji želimo predvidjeti te provodimo diskretizaciju numeričkih

vrijednosti u 13 razreda. Zatim se provodi prvi podproces u kojemu pomoću Rule Induction operatora treniramo model i kreiramo pravila. Ta se pravila zatim na test setu testiraju i mjerimo performanse modela. Zatim provodimo unakrsnu validaciju modela. Kao i u prošlom modelu, provodi se 10 validacija modela. Zatim učitalamo netestiranu bazu podataka te ponavljamo postupak do operatora za primjenu modela kako je prikazano na slici 41. Nakon toga očitavamo rezultate provedene analize.

Na sljedećim slikama vidimo kolika je preciznost modela. Postignuta preciznost iznosi 66,19%, što je dobar rezultat. Budući da postoji 13 klasa, kreiran je veliki broj poslovnih pravila od kojih je za set podataka bilo 1720 točnih od 2310 prema generiranim poslovnim pravilima.

accuracy: 66.19% +/- 3.97% (mikro: 66.19%)						
	true Velika B	true Češka	true Ukraine	true Poljska	true Italija	true Hrvatska
pred. Velika	642	122	93	19	6	23
pred. Češka	63	162	15	16	0	7
pred. Ukrain	3	3	74	1	0	2
pred. Poljska	5	4	1	209	0	0
pred. Italija	0	0	0	0	8	0
pred. Hrvatska	1	0	0	0	0	26
pred. Njema	10	18	7	3	1	3
pred. Francu	0	0	0	0	0	0
pred. Ruska	3	13	12	4	7	4
pred. Sloven	1	4	1	1	0	0
pred. UNITE	0	0	0	0	0	0
pred. GIBRA	0	0	0	0	0	1
pred. Turska	0	0	0	0	0	0
pred. Belgija	0	0	0	0	0	0
class recall	88.19%	49.69%	36.45%	82.61%	36.36%	39.39%

Slika 43. Vektor performansi modela (1)



	true Njemač	true Francus	true Ruska F	true Slovenij	true UNITED	true GIBRAL	true Turska	true Belgija
	96	14	41	10	0	1	0	0
	34	4	23	19	6	5	0	4
	3	0	6	0	0	0	0	0
	0	0	4	0	0	0	0	0
	0	0	0	0	0	0	0	0
	2	0	1	0	0	2	0	0
	88	4	4	3	0	0	0	2
	0	3	0	0	0	0	0	0
	2	2	286	1	0	0	8	0
	0	1	2	26	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	5	0
	0	0	0	0	0	0	0	0
	39.11%	10.71%	77.93%	44.07%	0.00%	0.00%	38.46%	0.00%

Slika 44. Vektor performansi modela (2)

Sa vektora performansi moguće je uočiti jednu zanimljivu činjenicu. Vidi se kako se najveća preciznost postiže za one države isporuke gdje postoji veći broj isporuka. Tako se vidi kako je točnost modela za predviđanje isporuke u Veliku Britaniju 88,19%, a također vidimo kako je i najveći broj isporuka također u Velikoj Britaniji. Slična je situacija i sa Poljskom i Rusijom, gdje je postignuta veća točnost modela uz veći broj podataka koje je bilo moguće testirati. Model nije uspio predvidjeti isporuke u sljedeće države:

- Ujedinjeni Arapski Emirati
- Belgija
- Gibraltar

Ono što povezuje ove tri države isporuke je malen uzorak, odnosno malena mogućnost testiranja podataka i kreiranja poslovnih pravila koja će točno predvidjeti isporuku u navedene države. Na slici 45 prikazana su neka od generiranih poslovnih pravila.

```

if SumaOperacijaPodnalog = range10 [31.846 - 35.385] then Hrvatska (0 / 0 / 0 / 0 / 0 / 2 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)
if SumaOperacijaPodnalog = range9 [28.308 - 31.846] and SumaMinutaNalog = range1 [-∞ - 14133.731] then Ukraine (0 / 0
if SumaOperacijaNalog = range7 [142.615 - 166.385] and SumaOperacijaPodnalog = range4 [10.615 - 14.154] then Slovenija
if SumaMinutaPodnalog = range2 [3886.923 - 7738.846] then Francuska (0 / 0 / 0 / 0 / 0 / 0 / 0 / 5 / 0 / 0 / 0 / 0 / 0 / 0 / 0)
if SumaOperacijaPodnalog = range12 [38.923 - 42.462] and LANSKOL = range1 [-∞ - 39.385] then Hrvatska (0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)
if BrojPodnaloga = range3 [4.538 - 6.308] then Ukraine (0 / 0 / 1 / 0 / 0 / 0 / 1 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)
if SumaOperacijaNalog = range7 [142.615 - 166.385] and LANSKOL = range1 [-∞ - 39.385] then Poljska (0 / 3 / 4 / 4 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)
if BrojPodnaloga = range2 [2.769 - 4.538] and SumaMinutaNalog = range1 [-∞ - 14133.731] then Njemačka (1 / 2 / 1 / 1 / 1 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)
if SumaMinutaNalog = range1 [-∞ - 14133.731] and FaktorCijene = range3 [1.946 - 2.917] then Hrvatska (0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 4)
if LANSKOL = range3 [77.769 - 116.154] then Belgija (0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 4)
if SumaOperacijaNalog = range3 [47.538 - 71.308] and SumaOperacijaPodnalog = range4 [10.615 - 14.154] then Hrvatska (0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)
if BrojPodnaloga = range7 [11.615 - 13.385] then Italija (0 / 0 / 0 / 0 / 0 / 3 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)
if SumaMinutaNalog = range2 [14133.731 - 28267.462] then Velika Britanija (2 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)
if BrojPodnaloga = range2 [2.769 - 4.538] then Francuska (0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 2 / 0 / 0 / 0 / 0 / 0 / 0)
if SumaOperacijaNalog = range3 [47.538 - 71.308] and LANSKOL = range1 [-∞ - 39.385] then Hrvatska (1 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)
if LANSKOL = range1 [-∞ - 39.385] then Njemačka (0 / 0 / 1 / 1 / 1 / 0 / 1 / 1 / 1 / 0 / 0 / 0 / 0 / 0 / 0 / 0)

```

correct: 1720 out of 2310 training examples.

### Slika 45. Generirana poslovna pravila

Prikazana su modelirana poslovna pravila.

```

if VrijemeDoIsporuke = range13 [242.231 - ∞] then Velika Britanija (11 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)

```

```

if VrijemeDoIsporuke = range4 [-16.692 - 12.077] and SumaOperacijaNalog = range3 [47.538 - 71.308] then Velika Britanija (17 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)

```

```

if VrijemeDoIsporuke = range4 [-16.692 - 12.077] and BrojPodnaloga = range7 [11.615 - 13.385] then Velika Britanija (13 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)

```

```

if VrijemeDoIsporuke = range10 [155.923 - 184.692] and BrojPodnaloga = range6 [9.846 - 11.615] then Velika Britanija (20 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)

```

```

if SumaOperacijaNalog = range6 [118.846 - 142.615] and VrijemeDoIsporuke = range4 [-16.692 - 12.077] then Velika Britanija (12 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)

```

```

if SumaOperacijaNalog = range6 [118.846 - 142.615] and SumaMinutaNalog = range3 [28267.462 - 42401.192] then Velika Britanija (11 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)

```

```

if SumaOperacijaNalog = range6 [118.846 - 142.615] and SumaMinutaNalog = range4 [42401.192 - 56534.923] then Velika Britanija (8 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)

```

```

if SumaOperacijaNalog = range6 [118.846 - 142.615] and SumaOperacijaPodnalog = range11 [35.385 - 38.923] then Velika Britanija (8 / 0 / 2 / 0 / 0 / 0 / 0 / 2 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0)

```

```

if SumaOperacijaNalog = range6 [118.846 - 142.615] and SumaMinutaPodnalog = range3 [7738.846 - 11590.769] then Velika Britanija (6 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 3 / 0 / 0 / 0 / 0 / 0)

```

```

if SumaOperacijaNalog = range6 [118.846 - 142.615] and LANSKOL = range8 [269.692 - 308.077] then Velika Britanija (1 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 1 / 0 / 0 / 0 / 0 / 0)

```

```
if BrojPodnaloga = range3 [4.538 - 6.308] and SumaOperacijaNalog = range2
[23.769 - 47.538] then Velika Britanija (3 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0
/ 0 / 0 / 0 / 0 / 0)
```

```
if SumaOperacijaNalog = range6 [118.846 - 142.615] and BrojPodnaloga =
range3 [4.538 - 6.308] then Velika Britanija (24 / 0 / 13 / 0 / 0 / 0 / 0
/ 0 / 0 / 0 / 0 / 0 / 0 / 0)
```

```
if SumaOperacijaNalog = range6 [118.846 - 142.615] and BrojPodnaloga =
range4 [6.308 - 8.077] then Velika Britanija (43 / 7 / 0 / 0 / 0 / 0 / 10
/ 0 / 7 / 0 / 0 / 0 / 0 / 0)
```

```
if SumaOperacijaNalog = range13 [285.231 - ∞] and BrojPodnaloga = range11
[18.692 - 20.462] then Velika Britanija (18 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 /
0 / 0 / 0 / 0 / 0 / 0)
```

```
if VrijemeDoIsporuke = range4 [-16.692 - 12.077] and BrojPodnaloga = range3
[4.538 - 6.308] then Velika Britanija (16 / 0 / 0 / 0 / 0 / 0 / 0 / 5 / 0 / 0
/ 0 / 0 / 0 / 0 / 0)
```

```
if SumaOperacijaNalog = range5 [95.077 - 118.846] and BrojPodnaloga =
range5 [8.077 - 9.846] then Velika Britanija (14 / 0 / 0 / 0 / 0 / 0 / 0 / 0 /
0 / 0 / 0 / 0 / 0 / 0 / 0)
```

```
if BrojPodnaloga = range6 [9.846 - 11.615] and VrijemeDoIsporuke = range8
[98.385 - 127.154] then Velika Britanija (10 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 /
0 / 0 / 0 / 0 / 0 / 0)
```

```
if LANSKOL = range13 [461.615 - ∞] and BrojPodnaloga = range3 [4.538 -
6.308] then Velika Britanija (1 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 /
0 / 0 / 0)
```

```
if BrojPodnaloga = range3 [4.538 - 6.308] and LANSKOL = range4 [116.154 -
154.538] then Velika Britanija (1 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0
/ 0 / 0 / 0)
```

```
if BrojPodnaloga = range3 [4.538 - 6.308] and SumaMinutaPodnalog = range6
[19294.615 - 23146.538] then Velika Britanija (1 / 0 / 0 / 0 / 0 / 0 / 0 / 0 /
0 / 0 / 0 / 0 / 0 / 0 / 0)
```

```
if BrojPodnaloga = range6 [9.846 - 11.615] and SumaOperacijaPodnalog =
range7 [21.231 - 24.769] then Velika Britanija (9 / 2 / 2 / 0 / 0 / 0 / 0 / 0
/ 0 / 0 / 0 / 0 / 0 / 0 / 0)
```

```
if BrojPodnaloga = range6 [9.846 - 11.615] and SumaMinutaPodnalog = range3
[7738.846 - 11590.769] then Velika Britanija (2 / 1 / 0 / 0 / 0 / 0 / 0 / 0 /
0 / 0 / 0 / 0 / 0 / 0 / 0)
```

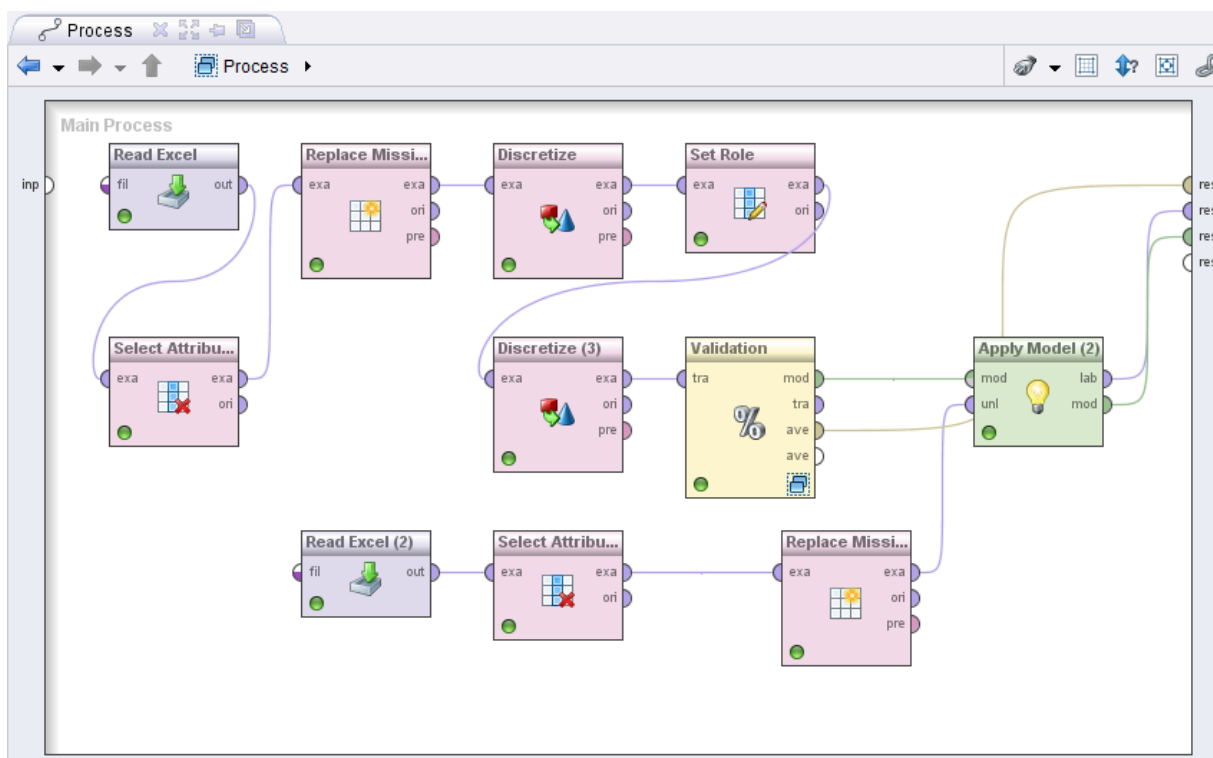
```
if BrojPodnaloga = range6 [9.846 - 11.615] and SumaOperacijaPodnalog =
range13 [42.462 - ∞] then Velika Britanija (1 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 0
/ 1 / 0 / 0 / 0 / 0 / 0 / 0)
```

```
if BrojPodnaloga = range6 [9.846 - 11.615] and SumaMinutaNalog = range1 [-∞
- 14133.731] then Velika Britanija (93 / 21 / 30 / 0 / 0 / 0 / 0 / 0 / 0 / 0 /
0 / 0 / 0 / 0 / 0)
```

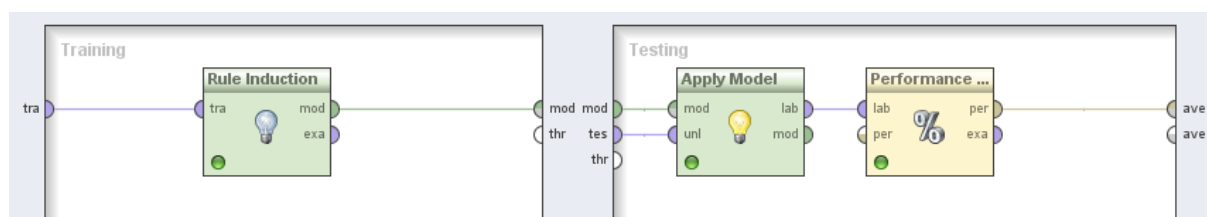
```
if BrojPodnaloga = range3 [4.538 - 6.308] and LANSKOL = range2 [39.385 -
77.769] then Velika Britanija (3 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 2 / 0 / 0 / 0 /
0 / 0 / 0)
```

### 5.3.4. Prediktivni model s induktivnim učenjem

U sljedećem modelu želimo predvidjeti faktor cijene, ali pomoću induktivnog učenja. Za ovaj primjer biti će korištena peta radna stranica baze podataka. U njoj postoji 573 redaka podataka, a za trening i test set biti će korišteno 500 podataka. Ostalih 73 podataka biti će iskorišteni za testiranje i validaciju modela. Svi operatori koje smo ranije koristili biti će potrebni i u ovoj analizi pa ih nije potrebno ponovno objašnjavati. Izradom ovog modela također će biti moguće i usporediti rezultate sa modelom za predviđanje faktora cijene pomoću neuronskih mreža. Bitno je napomenuti da algoritam ovdje radi sa razredima, i to sa svim atributima. To znači da predviđamo samo razred faktora cijene, ali ne i točnu vrijednost. Ovo ne predstavlja veliki problem zbog toga što je odabrano 11 razreda prema već prije spomenutoj formuli. Pomoću ovoga modela također želimo saznati hoće li manji broj podataka utjecati na točnost kreiranoga modela ili će i manji broj podataka biti dovoljan za kreiranje preciznoga modela. Na slikama 46 i 47 prikazani su glavni proces i dva potrebna podprocesa.



Slika 46. Glavni proces



Slika 47. Trening i test podproces

Sa slike 46 možemo primijetiti kako se model razlikuje od prijašnjih po tome što imamo dva operatora za diskretizaciju vrijednosti. Potrebna su nam dva ova operatora kako bi uspješno proveli diskretizaciju varijabli i kako bi operator za unakrsnu validaciju prihvatio razrede faktora cijene kao označene varijable za predviđanje. U test podprocesu ovoga puta korišten je operator za mjerenje performansi modela, ali klasifikacijski, što znači da može mjeriti preciznost previđanja klasa, što je upravo ono što nam je potrebno prilikom kreiranja ovakvog modela.

Nakon provođenja unakrsne validacije modela slijedi primjena testiranog modela na ostalih 73 podataka. Rezultati i preciznost modela prikazani su na sljedećim slikama.

accuracy: 75.76% +/- 2.83% (mikro: 75.75%)						
	true range1 [	true range2 [	true range3 [	true range4 [	true range5 [	true range6 [
pred. range1	347	38	1	1	5	2
pred. range2	13	19	3	2	0	1
pred. range3	1	4	6	6	4	1
pred. range4	0	1	6	5	1	1
pred. range5	1	5	4	7	1	0
pred. range6	0	0	0	0	0	0
pred. range7	0	0	0	0	1	0
pred. range8	0	0	0	0	0	0
pred. range9	0	0	0	0	0	0
pred. range1	0	0	0	0	0	0
pred. range1	0	0	0	0	0	0
class recall	95.86%	28.36%	30.00%	23.81%	8.33%	0.00%

Slika 48. Preciznost modela (1)

true range7 [	true range8 [	true range9 [	true range10	true range11	class precisi
0	0	4	2	0	86.75%
0	1	0	2	1	45.24%
0	0	0	0	0	27.27%
0	0	0	0	0	35.71%
2	0	0	0	0	5.00%
0	0	0	0	0	0.00%
0	0	0	0	0	0.00%
0	0	0	0	0	0.00%
0	0	0	0	0	0.00%
0	0	0	0	0	0.00%
0	0	0	0	0	0.00%
0.00%	0.00%	0.00%	0.00%	0.00%	

Slika 49. Preciznost modela (2)

Preciznost modela iznosi 75,76%, što predstavlja vrlo dobar rezultat s obzirom na ograničen broj podataka. Zbog manjeg broja podataka su bili kreirani i lošiji trening i test modeli, ali je model svejedno postigao vrlo dobru točnost. Može se zaključiti kako su diskretizirani podaci bolji za analizu induktivnim učenjem nego numerički učenjem neuronske mreže u slučaju kada imamo manji broj podataka za trening i test modela. Isto se može zaključiti i iz činjenice kako je najveći broj podataka koji su ušli u trening i test podprocese ušli u prvi razred, čija preciznost predviđanja iznosi čak 95,86%, za razliku od predviđanja klasa od 6 do 11, koji imaju točnost 0%. Ovo je posljedica toga što za te razrede induktivnim učenjem ne postoji dovoljan broj podataka čime se omogućilo bolje i preciznije kreiranje modela. Pomoću ovoga modela poduzeća odmah može reagirati prema potencijalnom kupcu zbog toga što temeljem kreiranih poslovnih pravila zna u kojem razredu će se kretati cijena proizvoda. Slijedi prikaz kreiranih pravila.

```

if SumaMin = range1 [-∞ - 16703.500] then range1 [-∞ - 1.153]

if SumaMin = range3 [33407 - 50110.500] and BrojPodnaloga = range2 [3.091 - 5.182] then range2 [1.153 - 2.299]

if SumaMin = range2 [16703.500 - 33407] and BrojPodnaloga = range4 [7.273 - 9.364] then range2 [1.153 - 2.299]

if SumaMin = range2 [16703.500 - 33407] and VrijemeDoIsporuke = range4 [-1 - 33] then range1 [-∞ - 1.153]

if SumaMin = range3 [33407 - 50110.500] and SumaOperacija = range9 [224.727 - 252.818] then range2 [1.153 - 2.299]

```

```
if SumaMin = range2 [16703.500 - 33407] and BrojPodnaloga = range5 [9.364 - 11.455] then range2 [1.153 - 2.299]

if SumaMin = range3 [33407 - 50110.500] and VrijemeDoIsporuke = range5 [33 - 67] then range2 [1.153 - 2.299]

if SumaOperacija = range7 [168.545 - 196.636] and SumaMin = range6 [83517.500 - 100221] then range4 [3.446 - 4.592]

if SumaOperacija = range8 [196.636 - 224.727] and SumaMin = range5 [66814 - 83517.500] then range3 [2.299 - 3.446]

if SumaMin = range2 [16703.500 - 33407] and SumaOperacija = range5 [112.364 - 140.455] then range1 [-∞ - 1.153]

if SumaMin = range2 [16703.500 - 33407] and SumaOperacija = range6 [140.455 - 168.545] then range2 [1.153 - 2.299]

if BrojPodnaloga = range6 [11.455 - 13.545] and VrijemeDoIsporuke = range4 [-1 - 33] then range4 [3.446 - 4.592]

if SumaMin = range2 [16703.500 - 33407] and SumaOperacija = range2 [28.091 - 56.182] then range1 [-∞ - 1.153]

if SumaMin = range4 [50110.500 - 66814] and SumaOperacija = range6 [140.455 - 168.545] then range3 [2.299 - 3.446]

if SumaMin = range5 [66814 - 83517.500] and VrijemeDoIsporuke = range5 [33 - 67] then range4 [3.446 - 4.592]

if BrojPodnaloga = range3 [5.182 - 7.273] and VrijemeDoIsporuke = range4 [-1 - 33] then range2 [1.153 - 2.299]

if VrijemeDoIsporuke = range1 [-∞ - -69] then range2 [1.153 - 2.299]

if VrijemeDoIsporuke = range6 [67 - 101] and SumaMin = range4 [50110.500 - 66814] then range3 [2.299 - 3.446]

if SumaMin = range2 [16703.500 - 33407] and SumaOperacija = range8 [196.636 - 224.727] then range1 [-∞ - 1.153]

if VrijemeDoIsporuke = range4 [-1 - 33] and BrojPodnaloga = range11 [21.909 - ∞] then range4 [3.446 - 4.592]

if SumaMin = range2 [16703.500 - 33407] and BrojPodnaloga = range1 [-∞ - 3.091] then range2 [1.153 - 2.299]

if BrojPodnaloga = range3 [5.182 - 7.273] and VrijemeDoIsporuke = range6 [67 - 101] then range3 [2.299 - 3.446]

if SumaMin = range2 [16703.500 - 33407] and VrijemeDoIsporuke = range5 [33 - 67] then range1 [-∞ - 1.153]

if BrojPodnaloga = range6 [11.455 - 13.545] and VrijemeDoIsporuke = range7 [101 - 135] then range4 [3.446 - 4.592]

else range3 [2.299 - 3.446]
```

#### 5.4. Usporedba rezultata modela

Nakon dobivenih rezultata pomoću kojih se predviđa faktor cijene proizvoda slijedi usporedba modela. Korišteni su modeli s neuronskim mrežama i induktivnim učenjem. Oba modela imaju svoje prednosti i nedostatke. Neuronske mreže rade samo sa numeričkim vrijednostima tako da s njima nije moguće predvidjeti neke tekstualne podatke, poput države isporuke. Prednost neuronskih mreža vidi se kada radimo samo s numeričkim atributima zbog toga što ih nije potrebno diskretizirati. Diskretizacijom se zapravo gubi dosta veliki dio korelacije između ulaznih i izlaznih parametara kako se vidi sa slike 51. Na slici 51 prikazan je vektor performansi za induktivno učenje pomoću kojega se predviđa faktor cijene. Vidimo sa slike kao je koeficijent korelacije i determinacije dosta niži nego u slučaju modela sa neuronskim mrežama. S druge strane vidimo kako je postignuta točnost od 75%, ali to je u slučaju kada predviđamo samo razrede faktora cijene. Pomoću neuronskih mreža sa slike 50 koja predstavlja prije postignuti vektor performansi vidimo kako su koeficijenti korelacije i determinacije puno veći, što znači da model kreiran sa neuronskim mrežama bolje objašnjava skup podataka. Problem je u primjenjivosti neuronskih mreža. Neuronske mreže je bolje koristiti u slučajevima kada imamo numeričke podatke, a induktivno učenje kada moramo numeričke podatke diskretizirati.

```
PerformanceVector  
  
PerformanceVector:  
root_mean_squared_error: 0.956 +/- 0.185 (mikro: 0.974 +/- 0.000)  
normalized_absolute_error: 0.440 +/- 0.150 (mikro: 0.438)  
root_relative_squared_error: 0.565 +/- 0.072 (mikro: 0.572)  
squared_error: 0.949 +/- 0.375 (mikro: 0.949 +/- 4.779)  
correlation: 0.853 +/- 0.018 (mikro: 0.829)  
squared_correlation: 0.727 +/- 0.031 (mikro: 0.688)
```

Slika 50. Vektor performansi neuronskih mreža



## PerformanceVector

PerformanceVector:

accuracy: 75.76% +/- 2.83% (mikro: 75.75%)

ConfusionMatrix:

True:	range1 [-∞ - 1.153]	range2 [1.153 - 2.299]	range3		
range1 [-∞ - 1.153]:	347	38	1	1	5
range2 [1.153 - 2.299]:	13	19	3	2	0
range3 [2.299 - 3.446]:	1	4	6	6	4
range4 [3.446 - 4.592]:	0	1	6	5	1
range5 [4.592 - 5.739]:	1	5	4	7	1
range6 [5.739 - 6.886]:	0	0	0	0	0
range7 [6.886 - 8.032]:	0	0	0	0	1
range8 [8.032 - 9.179]:	0	0	0	0	0
range9 [9.179 - 10.325]:		0	0	0	0
range10 [10.325 - 11.472]:		0	0	0	0
range11 [11.472 - ∞]:	0	0	0	0	0

correlation: 0.490 +/- 0.096 (mikro: 0.464)  
squared\_correlation: 0.250 +/- 0.094 (mikro: 0.216)

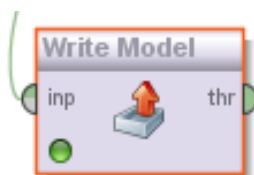
Slika 51. Vektor performansi modela za induktivno učenje

### 5.5. Zapisivanje i spremanje modela

Nakon kreiranja modela potrebno je spremiti model kako bi bio spreman za kasniju upotrebu na nekom drugom računalu gdje je prisutna veća baza podataka ili novija i osvježena baza podataka. Ovaj proces radimo pomoću operatora za zapisivanje modela (eng. *Write Model*). Moguće je model spremiti na tri načina:

- Kompresiranu XML datoteku
- XML datoteku
- Binarni oblik

RapidMiner programski paket omogućava otvaranje svih navedenih tipova podataka tako da je svejedno koji odaberemo sve dok je i na ostalim računalima prisutan isti programski paket. Operator za zapisivanje modela je prikazan na slici 52.



Slika 52. Operator za zapisivanje modela

## 6. ZAKLJUČAK

U radu su opisane osnovne značajke sustava poslovne inteligencije. Opisani su alati poslovne inteligencije i rudarenja podataka. U teoretskom djelu rada razrađene su mogućnosti i očekivani rezultati uvođenja sustava poslovne inteligencije u razna poduzeća. U radu su objašnjeni operatori programskog paketa RapidMiner koji je i poslužio kao podloga praktičnog djela rada.

U praktičnom djelu rada željelo se dokazati kako se pomoću alata poslovne inteligencije mogu kreirati modeli koji će poboljšati poslovanje maloserijskog proizvodnog poduzeća. Bilo je potrebno obraditi bazu podataka poduzeća koja je dobivena praćenjem proizvodnje pomoću informacijskog sustava. U radu su kreirana 4 modela, od kojih je jedan opisni, a ostali su prediktivni. Cilj opisnog modela bio je dokazivanje hipoteze kako se na istom radnom nalogu poduzeća nalaze pozicije za proizvodnju koje su neophodne jedna bez druge. Analiza je potvrdila ovu hipotezu, te je dokazano da pomoću modela za kreiranje asocijativnih pravila dobivamo upravo ciljane pozicije. Pomoću prediktivnih modela željelo se predvidjeti koliki će biti faktor cijene s obzirom na dostupne podatke iz baze. U analizu se krenulo s dva moguća modela. Prvi je bio pomoću neuronske mreže, a drugi pomoću induktivnog učenja. Neuronska mreža se pokazala kao bolja i preciznija opcija zbog veće fleksibilnosti operatora za učenje i testiranje neuronske mreže. Neuronska mreža se pokazala neupotrebljivom u slučaju kada se željelo predvidjeti državu isporuke pozicija. U ovom slučaju je kreiran model za induktivno učenje pomoću kojega su se numerički atributi u bazi diskretizirali, a zatim ušli u analizu. Analiza se pokazala kao vrlo dobra, a jedini nedostatak modela bio je manjak podataka za treniranje i testiranje modela. Također su i modelirana poslovna pravila pomoću induktivnog učenja za što nam je poslužio isti operator. Pomoću poslovnih pravila je moguće poboljšati organizaciju poduzeća, a pomoću prediktivnih modela moguće je bolje planirati proizvodnju s ciljem bržeg donošenja poslovnih odluka kada su u pitanju potencijalni kupci. Pomoću prediktivnih modela kupcu je moguće odmah reći kolika će biti cijena proizvodnje nekog proizvoda. Prediktivni modeli su zapisani i spremljeni kao XML datoteka čime je omogućeno njihovo olakšano korištenje na raznim mjestima. Usporedbom operatora za treniranje neuronske mreže i kreiranja poslovnih pravila diskretizacijom numeričkih varijabli zaključeno je kako su neuronske mreže preciznije od modela koji su koristili induktivno učenje. Nedostatak neuronskih mreža je nemogućnost rada

sa tekstualnim atributima za razliku od induktivnog učenja. U radu se vidi kako su neuronske mreže pogodnije za rad s numeričkim atributima zbog toga što kod njih nema potrebe za diskretizacijom numeričkih vrijednosti. Razlika se očitovala u puno većim koeficijentima korelacije i determinacije. Induktivno učenje je korisno kada želimo stvoriti poslovna pravila pomoću kojih želimo povećati kvalitetu organizacije poduzeća i postići automatizam u donošenju poslovnih odluka pomoću poslovnih pravila.

Poduzeća koja se bave maloserijskom proizvodnjom i imaju implementiran informacijski sustav pomoću kojega su u mogućnosti pratiti proizvodni proces generiraju veliku količinu podataka od koje nema velike koristi. Alatima poslovne inteligencije i rudarenja podataka mogu se izvući vrijedne informacije pomoću kojih je moguće predvidjeti budući trend na tržištu. Također je moguće i modeliranje poslovnih pravila, koja razvojem sustava umjetne inteligencije mogu automatski donositi važne poslovne odluke na temelju matematičkih modela i alata sustava poslovne inteligencije i rudarenja podataka.

## LITERATURA

- [1] [https://www.fsb.unizg.hr/atlantis/upload/newsboard/13\\_04\\_2012\\_\\_16748\\_K1\\_INFORMACIJSKI\\_MENADZMENT.pdf](https://www.fsb.unizg.hr/atlantis/upload/newsboard/13_04_2012__16748_K1_INFORMACIJSKI_MENADZMENT.pdf) - 27.10.2014.
- [2] Ken Withee - Microsoft business intelligence for dummies, 2010., Wiley
- [3] Business Intelligence – Data mining and optimization for decision making – Carlo Vercellis, John Wiley & Sons, 2009 Langit
- [4] Lynn: Smart Business Intelligence Solutions with Microsoft SQL Server 2008, Microsoft Press, 2009
- [5] Josip Kopejtko, Diplomski rad, 2012., Fakultet Strojarsva i Brodogradnje
- [6] Ivan Kramar, Diplomski rad, 2013., Fakultet Strojarsva i Brodogradnje
- [7] Ross, Ronald G. i Lam, Gladys S.W. Business Analysis with Business rules, Business rule Solutions, 2011.
- [8] Ernie Mazuin Mohd Yusof, Mohd Shahizan Othman, Yuhanis Omar, Ahmad Rizal Mohd Yusof – The Study of the application of business intelligence in manufacturing, 2013., IJCSI International Journal of Computer Science Issues, Vol. 10
- [9] Business intelligence in manufacturing – MAIA Intelligence white paper, MAIA Intelligence, 2009.
- [10] Chris Maxcer - How to get started with business intelligence in manufacturing, 2009., A SearchManufacturingERP.com e-knjiga
- [11] Markus Hoffman, Ralf Klinkenberg, RapidMiner – Data mining use cases and business analytics applications, 2014., CRC Press
- [12] Fareed Akthar, Caroline Hahne, RapidMiner 5 Operator Reference, 2012. -pristupljeno 19.11.2014
- [13] Informatički menadžment – Modeli rudarenja podataka, Davor Kolar, Marina Tošić, 2013. Fakultet Strojarsva i Brodogradnje

## PRILOZI

- Prilog 1. CD-R disc
- Prilog 2. Modelirana poslovna pravila
- Prilog 3. Baza podataka poduzeća
- Prilog 4. Trening set za neuronsku mrežu
- Prilog 5. Test set za neuronsku mrežu
- Prilog 6. Trening set za induktivno učenje za predviđanje faktora cijene
- Prilog 7. Test set za induktivno učenje za predviđanje faktora cijene
- Prilog 8. Test set za induktivno učenje za predviđanje države isporuke
- Prilog 9. Trening set za induktivno učenje za predviđanje države isporuke
- Prilog 10. Predprocesuirana baza podataka za asocijativna pravila
- Prilog 11. Model za kreiranje asocijativnih pravila
- Prilog 12. Model za predviđanje faktora cijene pomoću neuronskih mreža
- Prilog 13. Model za predviđanje faktora cijene pomoću induktivnog učenja
- Prilog 14. Model za predviđanje države isporuke pomoću induktivnog učenja