

■ Aplikativni sistemi odkrivanja zakonitosti v podatkih kot nov tip sistemov za podporo odločanju v informacijskih sistemih

Rok Rupnik, Marjan Krisper
Univerza v Ljubljani, Fakulteta za računalništvo in informatiko
Tržaška 25, 1000 Ljubljana
rok.rupnik@fri.uni-lj.si, marjan.krisper@fri.uni-lj.si

Povzetek

V sodobnih organizacijah dobiva podpora odločanju vse večji pomen, zato se na tem področju nenehno pojavljajo nove tehnologije. Ena izmed njih je odkrivanje zakonitosti in vzorcev v podatkih. Pojav vmesnikov in njihova standardizacija omogočajo razvoj aplikativnih sistemov odkrivanja zakonitosti v podatkih, ki predstavljajo nov tip sistemov za podporo odločanju. Prispevek v uvodu predstavlja področje odkrivanja zakonitosti v podatkih in standarde tega področja. V nadaljevanju predstavlja aplikativne sisteme odkrivanja zakonitosti v podatkih in aplikativni sistem Infolab-DMDSS kot primer tovrstnih sistemov, ki so predstavljeni kot nov pristop k uporabi metod odkrivanja zakonitosti v podatkih v okviru informacijskih sistemov. V zaključku prispevek opredeli še vlogo odkrivanja zakonitosti v podatkih z namenom prikazati doprinos in dodano vrednost, ki jo v informacijskih sistemih v okviru podpore odločanju predstavlja uporaba odkrivanja zakonitosti v podatkih.

Abstract

Data Mining Application Systems as a New Type of Decision Support Systems in Information Systems

Decision support is becoming very important in modern organizations and this fact constantly gives rise to a number of new technologies in this area. Data Mining is one of them: the one which enables the discovery and extraction of patterns from data. The emergence of data mining application interfaces and the efforts for their standardization enable the development of data mining application systems which use data mining methods and algorithms and represent a new type of decision support systems. In the first part of the article the area of data mining and data mining standards is introduced. In the second part of the article we discuss data mining application systems as a new approach in the usage of data mining in information systems. As a part of the discussion the DMDSS application system as an example of data mining application system is introduced. In the conclusion we point out the role of data mining in the area of decision support in information systems and how it affects the added value.

1 UVOD

Podpora odločanju dobiva vse večji pomen v sodobnih organizacijah. Poglavitni razlog za to je dejstvo, da postaja za sodobne organizacije poleg uspešnega nadziranja poslovanja vse pomembnejša opredelitev ustreznih strateških usmeritev in vizije. Podlaga za njihovo postopno opredelitev je kvalitetna podpora odločanju in visoka stopnja akumuliranega znanja.

Podjetja za potrebe analiziranja podatkov in podpore odločanju večinoma uporabljajo statistična orodja in orodja OLAP. Uporaba tovrstnih orodij oz. metod, na katerih temeljijo ta orodja, praviloma zahteva od analitika, da na začetku podrobno definira problem oz. da natančno ve, na katera vprašanja išče odgo-

vore. Ugotovimo lahko, da obe omenjeni področji temeljita na t. i. deduktivnem pristopu, katerega glavna značilnost je prav v postavitvi ustrezne hipoteze [JSR-73 2004; Hirji 2001]. Vse statistične metode ne temeljijo na preverjanju hipotez, nekatere se ukvarjajo tudi z analiziranjem podatkov. Slaba stran omenjenih metod je, da je uspeh odkritja pomembnih podatkov, vzorcev in trendov lahko odvisen od naključja oz. sreče. V primeru uporabe statističnih orodij in v okviru njih metod, ki temeljijo na postavitvi hipoteze, pomeni sreča postavitve ustreznih hipotez. V primeru uporabe OLAP orodij pa je sreča lahko v izboru ustreznih atributov za večdimenzionalne preglede in

izboru ustreznih vrednosti za filtriranje podatkov. V vsakem primeru pa velja, da je tudi za najboljše analitike pri analiziranju podatkov in odločanju težko upoštevati večje število atributov naenkrat.

Odkrivanje zakonitosti v podatkih (*Data Mining*) je področje, ki predstavlja eno od možnih »rešitev« predhodno omenjene problematike. Metode odkrivanja zakonitosti v podatkih omogočajo odkrivanje skritih vzorcev in trendov v podatkih [Westphal 1998]. Izvajanje analize podatkov z uporabo metod odkrivanja zakonitosti v podatkih predstavlja induktivni pristop analiziranja podatkov (JSR-73, 2004). Področje odkrivanja zakonitosti v podatkih, metode odkrivanja zakonitosti v podatkih in potenciali tega področja bodo predstavljeni v nadaljevanju članka.

Poslanstvo informacijskih sistemov je, med drugim, omogočiti odločitvene procese oz. podporo odločanju in odkrivanje znanja iz podatkovnih baz.¹ Obe omenjeni področji sta medsebojno odvisni. Odkrivanje znanja iz podatkovnih baz na eni strani omogoča akumuliranje znanja in posledično omogoča kvalitetnejši odločitveni proces, na drugi strani pa odločitve postavljajo smernice, ki vplivajo na cilje procesov odkrivanja znanja iz podatkovnih baz. Uporaba odkrivanja zakonitosti v podatkih kot pristop za odkrivanje znanja iz podatkovnih baz posledično pomeni semantično integracijo odkrivanja zakonitosti v podatkih v odločitvene procese oz. podporo odločanju. Zaradi naraščajočega pomena podpore odločanju se pojavljajo nove in nove tehnologije in področja, ki prehajajo po uveljavitvi na znanstvenem in strokovnem področju v komercialno uporabo in postajajo operativna. Odkrivanje zakonitosti v podatkih je primer področja, ki predstavlja komercialno uporabo metod strojnega učenja in je po razširitvi in uveljavitvi na znanstvenem področju doseglo visoko stopnjo operativnosti.

1.1 Namen prispevka

Namen prispevka je predstaviti aplikativne sisteme odkrivanja zakonitosti v podatkih kot nov tip sistemov za podporo odločanju. Aplikativni sistem odkrivanja zakonitosti v podatkih omogočajo nov, drugačen način uporabe odkrivanja zakonitosti v podatkih za potrebe podpore odločanju. Predstavitev aplikativnih sistemov odkrivanja zakonitosti v podatkih bo opravljena v sklopih, ki razširjajo namen prispevka, in sicer:

• predstavitev standardov področja odkrivanja zakonitosti v podatkih s poudarkom na predstavitvi aplikacijskih vmesnikov kot gradnikov, ki omogočajo razvoj sistemov za podporo odločanju, ki temeljijo na uporabi metod odkrivanja zakonitosti v podatkih;

• predstavitev aplikativnega sistema Infolab-DMDSS (*Data Mining Decision Support System*) kot primer aplikativnega sistema odkrivanja zakonitosti v podatkih. Predstavljene bodo funkcionalnosti sistema Infolab-DMDSS in metode odkrivanja zakonitosti v podatkih, ki jih podpira;

• predstavitev prednosti, ki jih predstavlja uporaba aplikativnih sistemov odkrivanja zakonitosti v podatkih v primerjavi z uporabo orodij za odkrivanje zakonitosti v podatkih;

• prikaz semantičnega prispevka k odločitvenim procesom, ki ga predstavlja uporaba odkrivanja zakonitosti v podatkih v informacijskem sistemu. V okviru tega je nakazan tudi semantični prispevek uporabe aplikativnih sistemov odkrivanja zakonitosti v podatkih.

2 ODKRIVANJE ZAKONITOSTI V PODATKIH

2 ODKRIVANJE ZAKONITOSTI V PODATKIH

Odkrivanje zakonitosti v podatkih (*Data Mining*) je področje, ki temelji na metodah in algoritmih umetne inteligence in njenega ožjega področja strojnega učenja [Kononenko 1997]. V slovenski strokovni javnosti se zanj uporabljajo še nekateri drugi izrazi, npr. podatkovno rudarjenje [Int01] in izkopavanje podatkov [Slovar]. Oba izraza ocenjujemo kot neprimerna, saj gre za bolj ali manj neposredne prevode angleškega izraza, ki bi ga že v osnovi lahko označili kot nekoliko ponesrečenega in nerodnega. Zato njegovo prevajanje ne more pripeljati do primernega izraza. Od pomensko ustreznih izrazov iskanje zakonitosti v podatkih in odkrivanje zakonitosti v podatkih ocenjujemo slednjega kot primernejšega, kar je posledica odtenka v pomenški razliki med izrazoma iskanje in odkrivanje. Pri iskanju gre praviloma za znan cilj, ki ga iščemo, ga želimo potrditi. Pri odkrivanju pa gre za neznano stvar (pojem, koncept, vzorec, pravilo), ki jo želimo odkriti, ugotoviti njen obstoj. Glede na naravo metod in algoritmov (predstavljeni bodo v nadaljevanju) tega

¹ Odkrivanje znanja iz podatkovnih baz je tukaj mišljeno kot splošno področje, kjer odkrivanje zakonitosti v podatkih predstavlja le enega od možnih pristopov.

področja zato ocenjujemo izraz odkrivanje zakonitosti v podatkih kot najprimernejši.

2.1 Kaj je Odkrivanje zakonitosti v podatkih

Različni avtorji podajajo svoje definicije področja odkrivanja zakonitosti v podatkih, vendar lahko ugotovimo, da so definicije podobne in se med seboj bistveno ne razlikujejo. V nadaljevanju podajamo nekatere od njih:

- Odkrivanje zakonitosti v podatkih je odkrivanje še neodkritih vzorcev in razmerij v podatkih. Gre za interdisciplinarno področje podatkovnih baz, strojnega učenja, odkrivanja vzorcev, statistike in vizualizacije [SolEuNet].
- Odkrivanje zakonitosti v podatkih je proces, v okviru katerega z uporabo različnih tehnik in orodij odkrivamo vzorce v podatkih. Gre za reševanje problemov z analiziranjem podatkov v podatkovni bazi [Moyle 2002].
- Odkrivanje zakonitosti v podatkih je odkrivanje skritih vzorcev in trendov v podatkih. Gre za razlaganje zakonitosti v podatkih in ne za testiranje hipotez [Westphal 1998].

Predstavljene definicije imajo semantično gledano visoko stopnjo preseka in vendarle vsaka svojo posebnost.

2.2 Pomembnejše metode odkrivanja zakonitosti v podatkih

Na področju iskanja zakonitosti v podatkih je znanih več metod. V nadaljevanju so na kratko predstavljene tiste, ki se pogosteje uporabljajo. Vsaka od njih se uporablja za različne namene in v različnih primerih, kar je posledica narave algoritmov, na katerih metode temeljijo.

2.2.1 Razvrščanje v skupine

Razvrščanje v skupine (*clustering*) je metoda, ki je uporabna v primerih, ko želimo ugotoviti tipične skupine primerov oz. porazdelitev posameznih primerov v določeno število naravnih skupin [Westphal 1998; Holsheimer 1998]. Skupine, ki predstavljajo rezultat uporabe metode, se med seboj razlikujejo, kar se kaže v različnih vrednosti pomembnejših atributov, znotraj skupine pa so si primeri podobni. Primer (element) je odvisen od problemske domene in je lahko oseba, subjekt, dogodek itn.

Metoda razvrščanja v skupine omogoča, da za posamezne probleme ugotovimo, katere so tipične skupine

s skupnimi značilnostmi, v katere lahko razdelijo množico primerov. Primera uporabe razvrščanja v skupine sta:

- prodajna organizacija lahko ugotovi, katere so tipične skupine njenih strank;
- državni organ lahko ugotovi tipične skupine državljanov, ki so oddali vlogo za izdajo gradbenega dovoljenja.

2.2.2 Asociacijska pravila

Asociacijska pravila (*association rules*) omogočajo identifikacijo razmerij, povezav, asociacij med vrednostmi oz. atributi v opazovani množici primerov [Westphal 1998; Holsheimer 1998]. Kot sinonim za asociacijska pravila se pojavlja tudi izraz povezovalna pravila, vendar menimo, da je izraz asociacijska pravila primernejši. Zelo znan in prvi večji primer uporabe asociacijskih pravil je analiza nakupovalne košarice, kjer so lastniki velikih trgovin z veliko prodajanimi artikli in veliko kupci želeli iz zapisov o posameznih nakupih izluščiti čimveč informacij o tem, kateri izdelki se najraje prodajajo skupaj. Na primer: če artikla A in B stranke velikokrat kupijo skupaj, potem ju je smotrno v trgovini postaviti enega zraven drugega, saj s tem še povečamo verjetnost, da kupec od nakupu enega kupi tudi drugega. V splošnem gre pri uporabi asociacijskih pravil za možnost odkrivanja zakonitosti, ki veljajo za primere, ki sodijo v izbrano problemsko množico. Povezovalna pravila so predstavljena na naslednji način:

$$X \rightarrow Y \text{ (zanesljivost, podpora)}$$

Podpora predstavlja pogostost nabora X in Y v transakcijah oz. dogodkih in je praviloma izražena relativno kot delež med vsemi transakcijami oz. dogodki. Zato podpora predstavlja tudi verjetnost, da se X in Y zgodita oz. nastopita hkrati. Zanesljivost pravila pa predstavlja pogojno verjetnost za Y pri pogoju X (verjetnost, da se ob dogodku X zgodi tudi dogodek Y).

Primer povezovalnega pravila iz področja nakupovalnih transakcij je naslednji:

$$\text{Teran} \rightarrow \text{Pršut} (0.83, 0.09)$$

Pravilo ima naslednji pomen: »Če je kupec kupil teran, je v 83 % primerov kupil tudi pršut. Delež kupcev, ki so kupili oba artikla, se je pojavil v 9 % vseh transakcij.«

2.2.3 Klasifikacija

Cilj pri reševanju klasifikacijskih problemov je identifikacija lastnosti razreda, v katerega sodijo primeri učne množice. Klasifikacija omogoča pridobitev modela lastnosti vnaprej predpisanih razredov. Tako pridobljeni model lastnosti ima dva pomena. Prvi pomen je v boljšem razumevanju primerov učne množice in s tem razreda. Drugi pomen pa je v možnosti napovedovanja, saj je z uporabo modela lastnosti možno napovedati razred, v katerega bodo razvrščeni novi primeri [Westphal 1998; Holsheimer 1998].

Klasifikacija ima dve fazi. V prvi fazi, fazi izdelave, se za določen razred nad primeri učne množice identificirajo njegove lastnosti. Rezultat prve faze je model lastnosti, ki je lahko predstavljen bodisi z odločitvenimi drevesi (*decision trees*) ali z množico klasifikacijskih pravil (*classification rules*). V drugi fazi, fazi uporabe, se pridobljeni model uporablja za klasifikacijo novih primerov v razrede. Za probleme, ki niso enkratnega značaja, je smotrno z neko smiselno frekvenco izdelovati novo učno množico in nov model lastnosti oz. ponavljati fazo izdelave, kar omogoča pridobivanje aktualnega modela in s tem boljšo uporabo modela v fazi uporabe [Westphal 1998].

Za klasifikacijo poznamo dve tehniki za predstavitev pravil oz. modela lastnosti razredov. Prva od njih je tehnika odločitvenih dreves. Klasifikacijska pravila so alternativa odločitvenim drevesom in predstavljajo drugo tehniko za predstavitev pravil. Cilj metode klasifikacijskih pravil je izpeljevanje množice if-then pravil za klasifikacijo primerov. Pravila imajo naslednjo strukturo: if pogoj then razred. Prvi del, pogoj, je sestavljen iz vrste testiranj oz. podpogojev, povezanih z logičnim operatorjem AND. Drugi del pravila pa poda razred, v katerega je primer klasificiran, lahko pa poda tudi verjetnostno porazdelitev po razredih. Primer klasifikacijskega pravila je:

IF (davčni zavezanec = ne) THEN stranka = SLABA
IF (davčni zavezanec = da) and (št. let partnerstva in (1,3)) THEN stranka = SREDNJE

Za metodo klasifikacije lahko podamo naslednji primer uporabe: podjetje želi ugotoviti, kakšne so lastnosti njenih dobrih, srednje dobrih in slabih strank. Postavi kriterije, ki opredeljujejo tri razrede strank: dobre, srednje dobre in slabe. Tipičen kriterij je lahko povprečna višina letnega prometa, kjer mora podjetje postaviti meje med razredi. Potem pripravi učno množico, kjer vsak zapis predstavlja eno stranko in v

kateri so najrazličnejši atributi, ki opisujejo stranko, poleg tega pa je v učni množici še klasifikacijski atribut, ki opredeljuje stranko: dobra, srednje dobra ali slaba. Potem lahko podjetje izdelava klasifikacijski model, s pomočjo katerega pridobi klasifikacijska pravila za vsakega od razredov strank. Od tega trenutka naprej lahko podjetje za vsako novo stranko ugotovi, v katero od treh skupin potencialno sodi. Vidimo, da je osnovna ideja klasifikacije izdelava modela lastnosti razredov, kar omogoča pridobitev znanja za neko problemsko domeno. Poleg tega pa je to znanje možno uporabiti nad množico novih primerov in ugotoviti, kam primer (potencialno) sodi.

2.3 Odkrivanje zakonitosti v podatkih in statistika

Področje odkrivanja zakonitosti v podatkih se ukvarja z odkrivanjem zakonitosti v podatkih, kar nakazuje na določeno sorodnost s področjem statistike. Predhodno smo že ugotovili, da ne temeljijo vse statistične metode na preverjanju hipotez, temveč se nekatere ukvarjajo tudi z analiziranjem podatkov. Za obe področji velja, da se ukvarjata z odkrivanjem zakonitosti v podatkih [Hand 1999]. Razlika med obema področjema je način, s katerim so predstavljeni modeli. Statistika v ta namen uporablja matematični jezik, odkrivanje zakonitosti v podatkih pa pravila, drevesa in druge načine, ki so praviloma bolj oz. lažje razumljivi uporabnikom, ki nimajo visokih znanj s področja statistike [Aggarwal 2002].

3 VPVELJAVA ODKRIVANJA ZAKONITOSTI V PODATKIH V INFORMACIJSKE SISTEME

Sodobna podjetja uporabljajo v okviru svojih informacijskih sistemov različne tipe aplikativnih sistemov za podporo odločanju [Laudon 1998]. Že nekateri transakcijski aplikativni sistemi vsebujejo kvalitetna poročila, izpise, katerih razvoj omogočajo najsodobnejša razvojna orodja. Prek konceptov sortiranja, grupiranja in statističnih operacij (sumiranje, minimum, maksimum, povprečje, standardni odklon) omogočajo ta poročila pridobivanje informacij, ki lahko predstavljajo podlago za odločanje na taktični ravni.

Direktorski aplikativni sistemi po mnenju nekaterih avtorjev ne sodijo med sisteme za podporo odločanju, saj ne upoštevajo znanja [Laudon 1998; Alter 2002]. Vendar je jasno, da tudi direktorski aplikativni sistemi predstavljajo obliko informacijske podpore odločitvenim procesom. Direktorski aplikativni sistemi praviloma temeljijo na podatkovnih skladiščih, ki se polnijo z

agregiranimi podatki iz več transakcijskih aplikativnih sistemov. Prek koncepta vrtanja v podatkih (*drill-down*) in koncepta več dimenzij omogočajo pregled nad (sumiranimi) podatki iz več zornih kotov, kjer dimenzije dejansko predstavljajo attribute.

Ugotovimo lahko, da za obe predstavljeni obliki informacijske podpore (poročila transakcijskih aplikativnih sistemov in direktorski aplikativni sistemi) za podporo odločanju velja, da gre za uporabo preprostih statističnih metod, operacij, kot so sumiranje, povprečje, maksimum itn. Predstavljene rešitve predstavljajo podlago za pridobivanje informacij in ne omogočajo odkrivanja vzorcev in razmerij med atributi. Podatki, pridobljeni prek preprostih statističnih operacij, omogočajo ob uporabi konceptov sortiranja in grupiranja le opazovanje razmerja večji/manjši v okviru enega atributa. Na primer: kdo ima večjo/manjšo realizacijo prodaje po regijah.

Podpora odločanju v okviru informacijskih sistemov doseže ob uporabi metod odkrivanja zakonitosti v podatkih višjo kakovostno raven, saj omogoča odkrivanje zakonitosti, ki veljajo med atributi in izdelavo modelov, ki predstavljajo pridobljeno znanje in omogočajo postopno akumuliranje znanja, kar predstavlja podlago za sprejemanje odločitev v prihodnosti.

3.1 Področja uporabe odkrivanja zakonitosti v podatkih v informacijskih sistemih

Uporaba metod odkrivanja zakonitosti v podatkih je v poslovnih sistemih v svetu že razširjena in se še vedno širi [Kohavi 2002]. Uporablja se na različnih področjih, predvsem področjih medicine, marketinga in zavarovalništva [Kukar 2002; Kononenko 2001; Kukar 1997; Grossman 2002; Apte 2002; Little 2002]. Nekatera okolja imajo z uporabo odkrivanja zakonitosti v podatkih zelo dobre izkušnje, saj so z njegovo uporabo uspešno razrešili problem ali dosegli zastavljeni cilj: izpeljali segmentacijo strank ali tržišča, odkrili in zmanjšali področja tveganja, izboljšali metode zdravljenja itn. [Kohavi 2002]. Področje, kjer v prihodnosti pričakujemo višjo stopnjo potrebe po uporabi odkrivanja zakonitosti v podatkih, je CRM (*Customer Relationship Management*), kjer so v okviru analitičnega CRM potrebne različne analize nad podatki o strankah [Rupnik 2001].

Višjo stopnjo uporabe metod odkrivanja zakonitosti v podatkih v informacijskih sistemih ovirajo naslednji problemi [Holsheimer 1998; Holsheimer 1999; Apte 2002]:

- nepoznavanje možnosti in potencialov, ki jih predstavlja področje odkrivanja zakonitosti v podatkih,
- nekatera orodja, ki omogočajo uporabo metod odkrivanja zakonitosti v podatkih, ne omogočajo uporabo metod odkrivanja zakonitosti v podatkih direktno nad podatki v podatkovni bazi, temveč zahtevajo pripravo podatkov v posebnem formatu, praviloma v tekstualnem,
- nekatera orodja ne omogočajo uporabe metod odkrivanja zakonitosti v podatkih nad večjimi količinami podatkov ali pa je v primeru večjih količin podatkov delovanje teh orodij nestabilno,
- orodja, ki omogočajo uporabo metod odkrivanja zakonitosti v podatkih neposredno nad podatki v podatkovni bazi, so bodisi zelo draga, bodisi ne podpirajo uporabe nad vsemi sistemi za upravljanje podatkovnih baz,
- učinkovita uporaba metod odkrivanja zakonitosti v podatkih zahteva sodelovanje vrhunskih strokovnjakov tega področja, saj gre za zelo kompleksne metode in pristope, ki ne omogočajo delovanja po nekaterih receptih,
- slabe možnosti uporabe metod odkrivanja zakonitosti v podatkih v odločitvenih sistemih in drugih zvrsteh aplikativnih sistemov oz. možnosti razvoja odločitvenih in aplikativnih sistemov z uporabo metod odkrivanja zakonitosti v podatkih.

Navedeni problemi povzročajo, da predstavlja v poslovnih okoljih in organizacijah uporaba odkrivanja zakonitosti v podatkih nemalokrat enkratne projekte in ne stalnega procesa. Koncept uporabe odkrivanja zakonitosti v podatkih kot enkratni ali občasni projekt sicer tudi lahko prinaša trenutne in kratkoročne rezultate, vendar prinaša dejanske in dolgoročne rezultate ter dodano vrednost samo ob stalni uporabi. Stalno uporabo narekuje tudi narava in kompleksnost metod odkrivanja zakonitosti v podatkih, saj se pri nemalo problemih kvalitetni podatkovni viri s kvalitetnimi atributi oblikujejo postopoma, v več iteracijah, kar zahteva določen čas [Westphal 1998]. Višjo stopnjo uporabe metod odkrivanja zakonitosti v podatkih v informacijskih sistemih ovira tudi pomanjkanje skupnih standardov na področju odkrivanja zakonitosti v podatkih.

3.2 Standardi področja odkrivanja zakonitosti v podatkih

Skupni standardi na področju odkrivanja zakonitosti v podatkih predstavljajo prvi potrebni korak oz. pogoj

za vpeljavo in širšo uveljavitev odkrivanja zakonitosti v podatkih v okviru informacijskih sistemov. Za izpolnitev omenjene vloge morajo standardi pokriti naslednja področja [Grossman 2002]:

- **modeli:** področje pokriva načine predstavitve modelov;
- **podatki:** področje pokriva pripravo, čiščenje, transformiranje in agregiranje podatkov pri izdelavi učnih množic oz. pripravi podatkovnih virov, ki predstavljajo podlago metodam odkrivanja zakonitosti v podatkih;
- **vmesniki:** področje pokriva aplikacijske vmesnike (API), ki omogočajo realizacijo in uporabo programskih knjižnic, ki omogočajo uporabo metod odkrivanja zakonitosti v podatkih v aplikacijah;
- **nastavitve:** področje pokriva parametre delovanja algoritmov, ki omogočajo uporabo metod odkrivanja zakonitosti v podatkih in njihove nastavitve;
- **proces:** področje pokriva proces uporabe metod odkrivanja zakonitosti v podatkih: priprave podatkovnih virov, izdelava modelov in uporaba modelov.

Glede na dosedanjo predstavitev področja standardov lahko tudi ugotovimo, da predstavljajo vmesniki enega ključnih elementov na področju standardov, saj pokrivajo izdelavo modelov, kar predstavlja centralni element uporabe metod odkrivanja zakonitosti v podatkih [Grossman 2002; JSR-73]. Cilj na področju standardizacije vmesnikov je narediti vmesnik, ki omogoča uporabo metod odkrivanja zakonitosti v podatkih, kar pomeni:

- možnost zamenjave inačice vmesnika, ne da bi bilo treba spreminjati aplikacijo,
- možnost izdelave modela za izbrano metodo direktno nad podatki v podatkovni bazi,
- možnost uporabe različnih algoritmov pri izdelavi modela,
- možnost specificiranja parametrov algoritmov in s tem možnost nadzora nad izdelavo modela,
- možnost shranjevanja modela v podatkovno bazo,
- možnost dostopa do modela in s tem možnost prikazovanja modela,
- možnost testiranja modela,
- možnost uporabe modela nad novimi podatki za metode napovedovalnega odkrivanja zakonitosti.

Doslej je znanih že nekaj primerov vmesnikov posameznih podjetij [Grossman 2002]. Oracle je v okviru sistema za upravljanje podatkovnih baz Oracle 9i prek posebne opcije omogočil uporabo metod od-

krivanja zakonitosti v podatkih neposredno nad podatki v podatkovni bazi Oracle [Int02]. V ta namen je podjetje razvilo tudi vmesnik Java API (OJDM), ki omogoča dostop do omenjene posebne opcije v podatkovni bazi in s tem uporabo metod odkrivanja zakonitosti v podatkih [Int03]. Vmesnik OJDM predstavlja možnost razvoja aplikacij, ki uporabljajo metode odkrivanja zakonitosti v podatkih. Gre za novo možnost, nov način uporabe metod odkrivanja zakonitosti v podatkih. Dosedanji način uporabe metod je temeljil na uporabi interaktivnih orodij, ki omogočajo ad hoc izdelavo modelov [Grossman 2002]. Tudi podjetje Microsoft je izdelalo vmesnik za potrebe izdelovanja modelov odkrivanja zakonitosti v podatkih za svoj sistem za upravljanje podatkovnih baz SQL Server 2000 [Grossman 2002].

Ugotovimo lahko, da predstavljajo različne realizacije nestandardiziranih vmesnikov oviro pri hitrem razširjanju vpeljave odkrivanja zakonitosti v podatkih v informacijske sisteme. Zato je v okviru Sun Java Community začel nastajati standardni vmesnik Java DataMining (JDM) kot Java Specification Request 73 (JSR-73) [Grossman 2002; JSR-73; Hornick 2003]. Cilj je določiti standardni vmesnik za okolji J2EE in J2SE, ki bo omogočal razvoj aplikacij v teh dveh okoljih, ki bodo omogočale uporabo metod odkrivanja zakonitosti v podatkih.

3.3 Metodologija uporabe odkrivanja zakonitosti v podatkih

Uporaba metod odkrivanja zakonitosti v podatkih je kompleksen postopek, ki zahteva metodološko podlago: opredelitev faz in opravil ter metodološke osnove. CRISP-DM (*CRoss Industry Standard Process for DataMining*) je metodologija, ki predstavlja procesni model, v okviru katerega opredeljuje faze in opravila, potrebne za izpeljavo projekta odkrivanja zakonitosti v podatkih [Crisp-dm]. Opredeljuje naslednjih šest faz:

- **razumevanje domene** – v okviru obravnavane faze je treba doseči razumevanje domene oz. problema, določiti cilje in zahteve projekta ter izdelati projektni plan;
- **razumevanje podatkov** – obravnavana faza razumevanje domene podkrepi z razumevanjem podatkovnih struktur, ki so na voljo za potrebe projekta;
- **priprava podatkov** – v okviru priprave podatkov se izvedejo priprava osnovnih podatkov, čiščenje,

transformiranje, agregiranje in formatiranje podatkov. Cilj faze je pripraviti podatke v obliko, ki bo zagotavljala izdelavo uporabnih modelov;

- **izdelava modelov** – v okviru izdelave modela je treba izbrati metode za izdelavo modelov, za izbrane metode izbrati najprimernejše algoritme, določiti vrednost parametrov algoritmov ter izdelati modele. Že v okviru izdelave modelov je prek metod potrebno izvajati ocenjevanja atributov in modelov ter glede na dobljene rezultate opraviti dodatne posege na pripravljenih podatkih;
- **evalvacija modelov** – evalvacija modelov mora zagotoviti izbor najprimernejših modelov glede na kriterije ocenjevanja modelov za različne metode. Glede na dobljene rezultate je treba določiti nadaljnje potrebne korake. Možnosti je več: od potrditve izbranih modelov za potrebe uporabe do vrnitve v fazo razumevanja domene;
- **uporaba modelov** – v okviru uporabe modela je treba najprej izdelati plan in opredeliti načine uporabe modelov. Poleg tega je treba opredeliti plan nadziranja uporabe modelov in njihovega vzdrževanja.

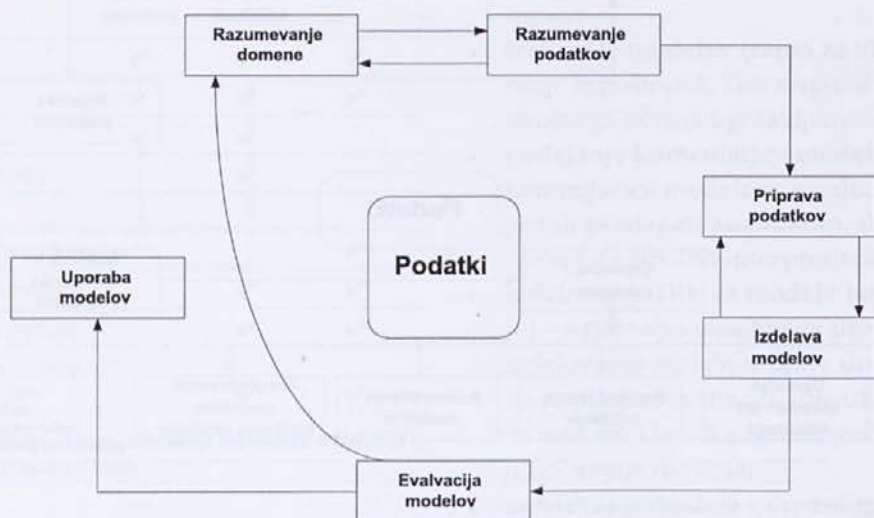
Kratka predstavitev metodologije nakazuje njeno projektno usmerjenost, kar načelno ne pogojuje stalnosti procesa (slika 1). V metodologiji pa so elementi, ki nakazujejo potrebo po stalnosti procesa uporabe

odkrivanja zakonitosti v podatkih. Faza uporabe modelov npr. obravnava nadziranje uporabe modelov in njihovega vzdrževanja. Ugotovimo lahko, da metodologija podpira tudi uporabo odkrivanja zakonitosti v podatkih kot stalni proces.

3.4 Aplikativni sistem INFOLAB-DMDSS

V letu 2002 smo za znanega naročnika izvedli študijo, katere cilj je bil ugotoviti potrebe in možnosti za področje podpore odločanju. V okviru študije je bilo ugotovljeno, da je za naročnika tehnologija odkrivanja zakonitosti v podatkih zelo primerna. Ob že omenjeni pojavitvi vmesnika Java API v sistemu za upravljanje podatkovnih baz Oracle 9i je bila izdelana dodatna študija, katere cilj je bil ugotoviti primernost uvajanja odkrivanja zakonitosti v podatkih prek razvoja in uporabe aplikativnega sistema odkrivanja zakonitosti v podatkih, ki temelji na vmesniku Java API. Aplikativni sistem odkrivanja zakonitosti v podatkih je sistem za podporo odločanju, ki temelji na uporabi metod odkrivanja zakonitosti v podatkih. Študija je dala pozitiven rezultat in naročnik je naročil razvoj aplikativnega sistema Infolab-DMDSS.

Namen pričujočega razdelka je predstaviti aplikativni sistem Infoab-DMDSS. V okviru predstavitve bodo predstavljene njegove funkcionalnosti in metode odkrivanja zakonitosti v podatkih, ki jih podpira. Cilj predstavitve je opozoriti na prednosti, ki jih



Slika 1: Shema metodologije CRISP-DM²

² Povzeto po [Crisp-dm].

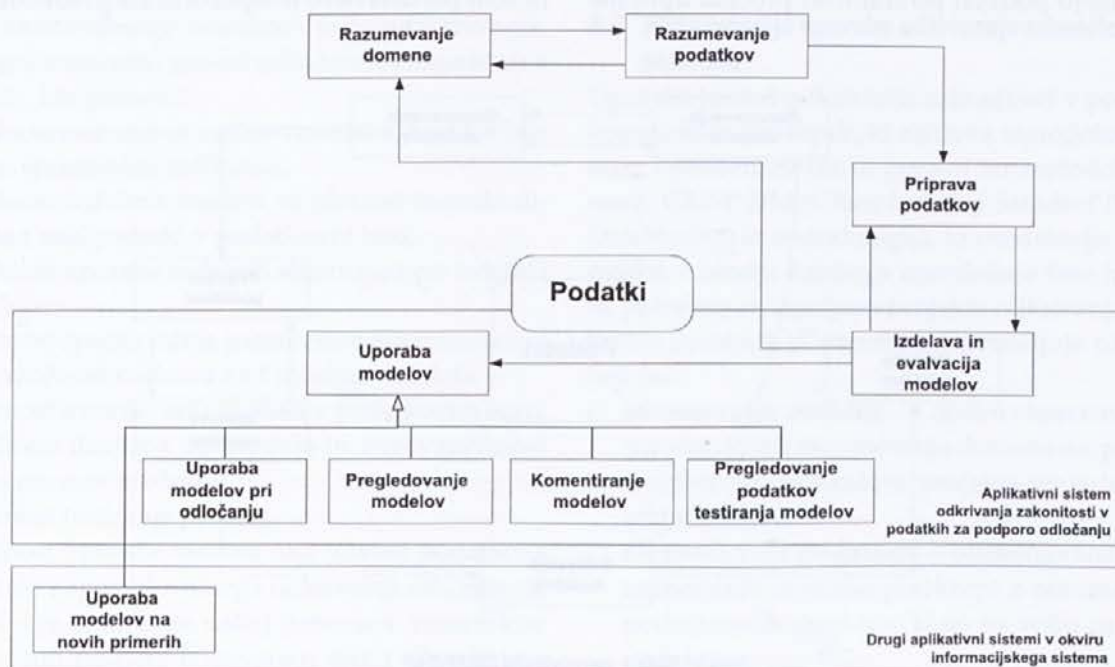
predstavlja uvajanje odkrivanja zakonitosti v podatkih prek aplikativnih sistemov za odkrivanje zakonitosti v podatkih.

3.4.1 Metodologija uporabe odkrivanja zakonitosti v podatkih za uporabo sistema Infolab-DMDSS

Pri zasnovi metodologije uporabe odkrivanja zakonitosti v podatkih za uporabo sistema Infolab-DMDSS smo temeljili na metodologiji CRISP-DM. Ocenili smo jo kot primerno, ugotovili pa smo, da je treba združiti fazi izdelave modela in evalvacije modela v eno, združeno fazo. Slika 2 prikazuje metodologijo uporabe metod odkrivanja zakonitosti v podatkih v informacijskem sistemu za potrebe sistema Infolab-DMDSS, kjer je poudarek na uporabi odkrivanja zakonitosti v podatkih kot stalnem procesu v primeru uporabe aplikacije odkrivanja zakonitosti v podatkih. Uvajanje odkrivanja zakonitosti v podatkih v informacijske sisteme prek aplikativnih sistemov odkrivanja zakonitosti v podatkih namreč temelji na oceni, da lahko da uporaba odkrivanja zakonitosti v podatkih dolgoročno gledano dobre rezultate le v primeru, ko gre za stalni proces [Holsheimer 1998; Kohavi 2002; Grossman 2002].

H glede na CRISP-DM nekoliko spremenjeni metodologiji uporabe odkrivanja zakonitosti v podatkih lahko podamo naslednje komentarje:

- Pri fazah razumevanja domene, razumevanja podatkov in priprave podatkov semantično ne gre za spremembe. Ker gre za aplikacijo, lahko v tem primeru govorimo o več domenah, ki jih v okviru aplikacije lahko poimenujemo analize. V okviru analiz lahko določamo podanalize, kjer analize izvajamo na podmnožicah.
- Za fazi izdelovanja modelov in evalvacija modelov ni več potrebno, da sta ločeni, saj se modeli izdelujejo in evalvirajo interaktivno, logično gledano gre za skupno fazo. Dobri modeli so označeni s posebnim statusom, kar omogoča njihovo uporabo oz. pregledovanje s strani analitikov.
- V okviru faze uporabe modelov lahko govorimo o več oblikah uporabe modelov, bodisi prek aplikativnega sistema, bodisi v okviru drugih aplikativnih sistemov v okviru informacijskega sistema. Prek aplikativnega sistema lahko uporabniki pregledujejo modele, dajejo modelom in posameznim pravilom v okviru modelov komentarje ter pregledujejo podatke o testiranju modelov [Aggrawal 2002]. Poleg tega lahko pri analizah področja klasifikacije klasifikacijska pravila uporabljajo tudi drugi aplikativni sistemi v okviru informacijskega sistema. Aplikativni sistem za področje prodaje lahko npr. že ob vnosu nove stranke prek ustreznega modela in njegovih



Slika 2: Metodologija uporabe metod odkrivanja zakonitosti v podatkih za Infolab-DMDSS

pravil ugotovi, v kateri razred nova stranka potencialno sodi glede na svoje lastnosti. Najpomembnejši način uporabe modelov pa je vsekakor njihova uporaba pri odločitvenih procesih.

Podane ugotovitve nakazujejo, da lahko govorimo pri aplikativnem sistemu odkrivanja zakonitosti v podatkih o treh glavnih vlogah njegovih uporabnikov, ki morajo medsebojno dobro sodelovati. Prva je vloga skrbnika podatkov, ki je zadolžen za fazo priprave podatkov in pri tem sodeluje s skrbniki vseh aplikativnih sistemov, ki predstavljajo vir podatkov za metode odkrivanja zakonitosti v podatkih. Druga je vloga skrbnika modelov, ki je zadolžen za fazo izdelave in evalvacije modelov. Ravno tako pa sodeluje pri komentiranju modelov, saj lahko tako analitikom daje opozorila, vezana na pridobljene modele in jim tako omogoči pravilno razumevanje in interpretiranje modela. Pomembna odgovornost skrbnika modelov je tudi v tem, da potrdi za fazo uporabe samo tiste modele, ki glede na kriterije ocenjevanja in evalviranja modelov presežejo določen prag, kar zagotavlja uporabnost modela. Tretja pa je vloga analitika, ki določa potrebe po novih problemskih domenah (analizah), pregleduje modele ter opredeljuje načine njihove uporabe. Analitik mora dobro poznati problemsko domeno, poznati pa mora tudi metode odkrivanja zakonitosti v podatkih. Tabela 1 prikazuje sodelovanje vseh treh vlog v okviru faz.

Faza/vloga	Skrbnik podatkov	Skrbnik modelov	Analitik
Razumevanje domene	✓	✓	✓
Razumevanje podatkov	✓	✓	✓
Priprava podatkov	✓	✓	
Izdelava in evalvacija modelov		✓	
Uporaba modelov			
Uporaba modelov pri odločanju			✓
Pregledovanje modelov		✓	✓
Komentiranje modelov		✓	✓
Pregledovanje podatkov testiranja modelov		✓	✓

Tabela 1: Vloge aplikativnega sistema odkrivanja zakonitosti v podatkih in njihovo sodelovanje v posameznih fazah

3.4.2 Funkcionalnosti sistema Infolab-DMDSS

Aplikativni sistem Infolab-DMDSS ima funkcionalnosti razdeljene glede na metodo odkrivanja zakoni-

nosti v podatkih in glede na vlogo. Glede na vlogo govorimo o funkcionalnostih, namenjenih analitiku, in funkcionalnostih, namenjenih skrbniku modelov. Glede na metodo odkrivanja zakonitosti v podatkih pa govorimo o funkcionalnostih za področja klasifikacije, asociacijskih pravil in razvrščanja v skupine.

Pri delitvi funkcionalnosti glede na vlogo je bila osnovna usmeritev ločiti fazo uporabe modelov od vseh drugih faz predhodno predstavljene metodologije uporabe odkrivanja zakonitosti v podatkih. Takšna ločitev je omogočila tudi povsem različen nivo potrebnih znanj za področje metod odkrivanja zakonitosti v podatkih za uporabo Infolab-DMDSS. Analitik tako potrebuje le osnovni nivo znanj, ki so potrebna za ustrezno interpretiranje modelov. Skrbnik modelov pa mora imeti visok nivo znanj za področje odkrivanja zakonitosti v podatkih: biti mora dober poznavalec metod in njihovih parametrov, znati mora evalvirati izdelane modele. Aplikativni sistem Infolab-DMDSS omogoča za vlogo skrbnika modelov naslednje funkcionalnosti:

- **izdelava modelov** – za vse tri predhodno navedene metode odkrivanja zakonitosti v podatkih omogoča izdelavo modelov. Pri izdelavi modelov je poleg vnosa namena izdelave modela mogoče izbrati tudi enega ali več parametrov, ki vplivajo na rezultat. Učne množice oz. podatkovni viri za izdelavo se avtomatsko osvežujejo vsako noč, zato je smiselno, da skrbnik podatkov za vsako analizo izdela nov model vsaj dvakrat do štirikrat na mesec;
- **testiranje modelov (samo za klasifikacijo)** – testiranje je postopek, kjer z uporabo priznanih metod strojnega učenja ugotavljamo kvaliteto modela;
- **podajanje komentarjev modelom** – cilj podajanja komentarjev modelom je z dodatnimi opisi in pojasnili pomagati analitikom, da bodo lažje, predvsem pa ustrezno interpretirali model;
- **izdelovanje slik za modele (samo za klasifikacijo)** – s pomočjo posebnega programa, ki omogoča izdelovanje različnih tipov drevesnih diagramov na podlagi ustrezne vhodne ukazne datoteke, smo za metodo klasifikacije omogočili izdelavo slike za odločitvena drevesa;
- **potrjevanje modela** – s potrditvijo modela skrbnik modela potrdi, da je model dovolj kvaliteten in primeren za objavo. Objavljeni modeli so tisti, ki so na voljo za uporabo analitikom. Ob doseganju ustreznega nivoja kvalitete modela je smiseln kriterij

za primernost modela to, da je različen od predhodno objavljenega modela opazovane analize;

- **evalvacija modela** – evalvacijo modela izvaja skrbnik modelov prek posebne forme, ki omogoča dostop do vseh navedenih in še naslednjih funkcionalnosti: pregledovanje modelov, pregledovanje podatkov o testiranju in pregledovanje slik za modele.

Za vlogo analitika pa omogoča naslednje funkcionalnosti:

- **pregledovanje modela** omogoča pregledovanje osnovnih podatkov o modelu (datum izdelave modela, namen izdelave modela idr.) in pravil modela. Pravila so prikazana v seznamu prek uporabe več prijmov za doseganje vizualno prijazne oblike: zamiki, krepki tisk za rezervirane besede ipd. Parametri delovanja aplikacije omogočajo spreminjanje rezerviranih besed oz. spreminjanje prevodov rezerviranih besed (npr. *if prevedemo lahko: »če«, »v kolikor«, »v primeru, da«*);
- **pregledovanje slik za modele (samo za klasifikacijo)** omogoča pregledovanje odločitvenih dreves. Gre za vizualizacijo informacije, ki je sicer že predstavljena v klasifikacijskih pravilih. Namen vizualizacije je še dodatno povečati razumljivost modela, ga približati analitiku in mu s tem omogočiti lažjo interpretacijo;
- **pregledovanje komentarjev za modele** – cilj komentarjev je analitikom omogočiti lažjo, predvsem pa ustrezno interpretacijo modela.

Vidimo, da so določene funkcionalnosti na voljo tako skrbniku modelov kot tudi analitiku. Nekatere od teh skupnih funkcionalnosti so za obe vlogi povsem enake, druge pa se glede na vlogo malenkostno razlikujejo.

3.4.3 Primer analize z uporabo aplikativnega sistema Infolab-DMDSS

V aplikativnem sistemu Infolab-DMDSS poteka uporaba metod odkrivanja zakonitosti v podatkih prek analiz, ki temeljijo na metodah odkrivanja zakonitosti v podatkih. Tipičen primer analize, primeren za večino organizacij, je klasifikacija strank. Klasifikacija strank omogoča pridobitev klasifikacijskih pravil, ki za posamezne vnaprej definirane razrede strank poda opise lastnosti posameznih razredov.

Za potrebe analize je bila po fazi razumevanja domene in v okviru faz razumevanja podatkov in priprave podatkov v več iteracijah definirana in postopno oblikovana učna množica. Za klasifikacijski atribut so bili definirani trije razredi: slab kupec, srednje dober kupec in dober kupec. Prek več iteracij faze izdelave modela in evalvacije modela je bil določen nabor parametrov algoritma in njihovih vrednosti. Za nekatere parametre so bile določene fiksne vrednosti, za nekatere pa nabor vrednosti, med katerimi lahko izbira skrbnik modelov pri izdelavi modela.

Slika 3 prikazuje formo za pregledovanje modela in strani skrbnika modelov, ki je podobna formi za pregle-

Modeli za analizo *Klasifikacija strank*

Modeli	Podatki o modelu : Model_klasifikacije_strank_V06	Akcije
Model_klasifikacije_strank_V06 Model_klasifikacije_strank_V06	Datum modela : 21.11.2003 Število pravil modela : 9 Število komentarjev modela : 1 Število razredov : 3 Zadnji komentar modela : Evaluacija modela je dala dobre rezultate, zato je model vreden zaupanja!	KOMENTARJI SLIKE DREVES OSTALI PODATKI REZULTATI TESTIRANJA
UPRAVLJANJE Z MODELI NAZAJ NA ZACETEK	Izberi razred : <input type="text" value="Vsi razred"/>	
	Pravila za model : Model_klasifikacije_strank_V06 V primeru, da je DAVCNI_ZAVEZANEC v množici (DA) potem je --> Klasifikacija stranke enaka (Slab kupec) V primeru, da je DAVCNI_ZAVEZANEC v množici (NE) potem je --> Klasifikacija stranke enaka (Srednje dober kupec) V primeru, da je PRAVNI_STATUS v množici (Nepredeljeno) potem je --> Klasifikacija stranke enaka (Dober kupec) V primeru, da je PRAVNI_STATUS v množici (Pravna oseba, Samostojni podjetnik) potem je --> Klasifikacija stranke enaka (Slab kupec) V primeru, da je PRAVNI_STATUS v množici (Fizična oseba) potem je --> Klasifikacija stranke enaka (Srednje dober kupec)	

Slika 3: Forma za pregledovanje modela klasifikacijskih pravil v aplikativnem sistemu Infolab-DMDSS

dovanje modelov s strani analitika. Forma nakazuje na dejavnost analitika v okviru faze uporabe modela. Prikazani model je bil izdelan nad izmišljenimi podatki.

3.4.4 Prednosti uporabe aplikativnih sistemov odkrivanja zakonitosti v podatkih

Aplikativni sistemi odkrivanja zakonitosti v podatkih predstavljajo po našem mnenju najprimernejši način uvajanja odkrivanja zakonitosti v podatkih v informacijske sisteme, saj gre v tem primeru za višjo stopnjo integriranosti tako v informacijski sistem kot tudi v procese odločanja [Bayardo 2001; Holsheimer 1999; Heinrichs 2003]. Glavne prednosti tovrstnih aplikativnih sistemov so:

- omogočajo fleksibilnost pri oblikovanju funkcionalnosti aplikativnega sistema in različnih, uporabniku prilagojenih načinov prikaza modelov. Funkcionalnost, ki jo omogočajo, je npr. komentiranje modelov in pravil s strani skrbnikov modelov, kar analitikom omogoča pravilno razumevanje in interpretiranje modela;
- omogočajo stalno uporabo metod odkrivanja zakonitosti v podatkih in stalno izvajanje različnih, medsebojno povezanih analiz, ki temeljijo na metodah odkrivanja zakonitosti v podatkih;
- ločitev vlog skrbnika modelov in analitika omogoča porazdelitev potrebnih znanj o metodah odkrivanja zakonitosti v podatkih v manjši meri na analitika in večji na skrbnika modelov;
- prek aplikativnega sistema pridobljene modele lahko za potrebe napovedovanja uporabljajo tudi drugi aplikativni sistemi v okviru informacijskega sistema.

Iz navedenega sledi, da omogočajo aplikativni sistemi odkrivanja zakonitosti v podatkih stalno izvajanje analiz prek uporabe metod odkrivanja zakonitosti v podatkih na način, ki zahteva od analitikov le osnovno raven poznavanja in razumevanja metod odkrivanja zakonitosti v podatkih.

3.5 Vloga odkrivanja zakonitosti v podatkih v informacijskih sistemih

Dosedanja razprava je nakazala potrebo po opredelitvi vloge odkrivanja zakonitosti v podatkih v informacijskih sistemih. Osnovni cilj je ugotoviti in predstaviti prispevek, ki ga uporaba odkrivanja zakonitosti v podatkih predstavlja pri podpori odločanja kot pomembni obliki informacijske podpore v okviru in-

formacijskih sistemov. Opredelitev vloge odkrivanja zakonitosti v podatkih v ožjem ali širšem smislu je tudi cilj nekaterih večjih projektov [SolEuNet; Academy].

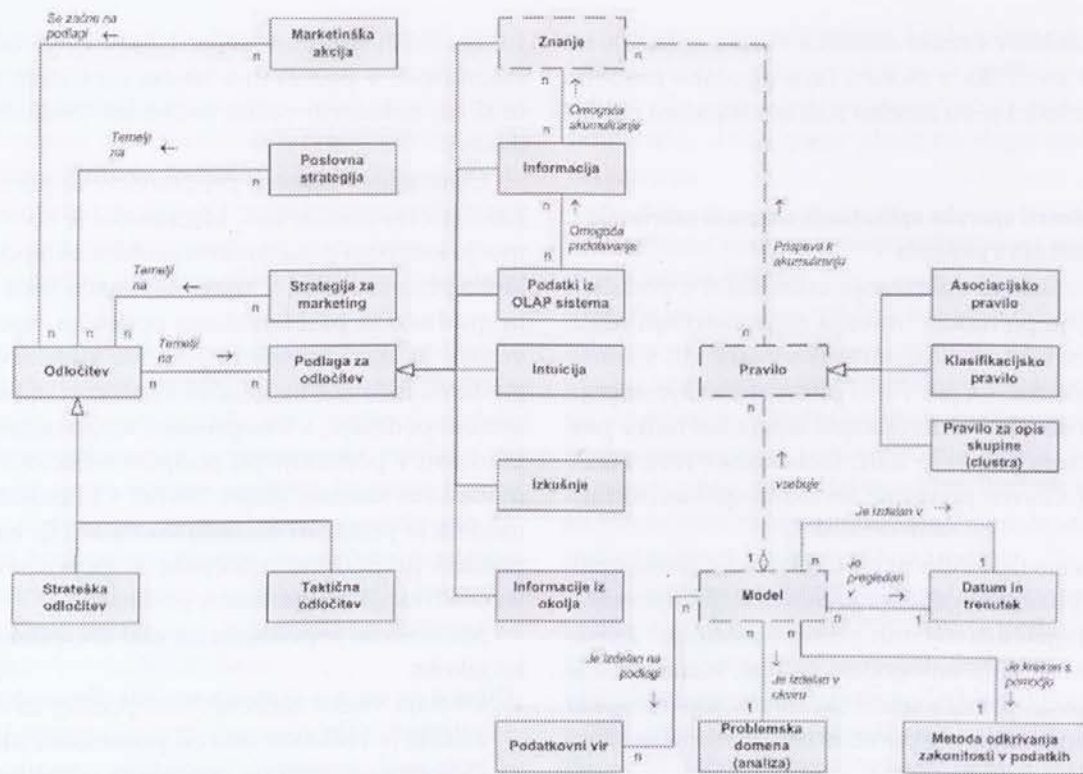
Opredelitev vloge in predstavitve doprinosa je prikazana z metamodelom. Metamodel je koncept, s pomočjo katerega predstavimo problemsko področje na konceptualni ravni. V njem nastopajo tako koncepti in gradniki iz problemskega področja, kot tudi povezave in razmerja med njimi. Na metamodel lahko gledamo tudi kot na miselni vzorec, ki opisuje problemsko področje, v tem primeru vlogo odkrivanja zakonitosti v podatkih pri podpori odločanju v informacijskem sistemu [Bajec 2001a]. Cilj izdelave metamodela je prikazati dodano vrednost, tj. kakovostni preskok pri podpori odločanju, ki ga prispeva uporaba odkrivanja zakonitosti v podatkih.

Metamodel je prikazan na sliki 4. Opišemo ga lahko takole:

- Obstaja veliko različnih vrst podlag za odločitve, znanje je vsekakor ena od pomembnejših.
- Odločitev temelji na podlagi za odločitev, na različnih odločitvah pa temeljijo različne strategije, med drugim poslovna in marketinška.
- Model vsebuje več pravil, vsako od njih pripada določenemu tipu pravil, ki je odvisen od metode odkrivanja zakonitosti v podatkih.
- Modeli oz. njihova pravila prispevajo k akumuliranju znanja, ki je verjetno najpomembnejša podlaga za odločanje.
- Model je izdelan na neki dan in v nekem trenutku (času) tega dneva.
- Analitik pregleda (uporabi) model na neki dan in v nekem trenutku (času) tega dneva.

Gradniki, prikazani s črtkasto ali pikčasto črto oz. robom, prikazujejo prispevek uporabe odkrivanja zakonitosti v podatkih podpori odločanju. Vidimo, da predstavljajo modeli in pravila novo kategorijo, ki prispevajo k akumuliranju znanja, ki predstavlja pomembno podlago za odločitve. Prispevek uporabe aplikativnih sistemov za odkrivanje zakonitosti v podatkih, je na metamodelu prikazana s pikčastimi črtami oz. robovi. Prispevek tovrstnih sistemov glede na klasični način uporabe odkrivanja zakonitosti v podatkih z uporabo orodij, je v večjih možnostih izdelave in uporabe modela v izbranem trenutku.

V kontekstu uporabe odkrivanja zakonitosti v podatkih v informacijskih sistemih lahko ugotovimo tudi, da je rezultat uporabe odkrivanja zakonitosti v



Slika 4: Metamodel vloge odkrivanja zakonitosti v podatkih v informacijskem sistemu

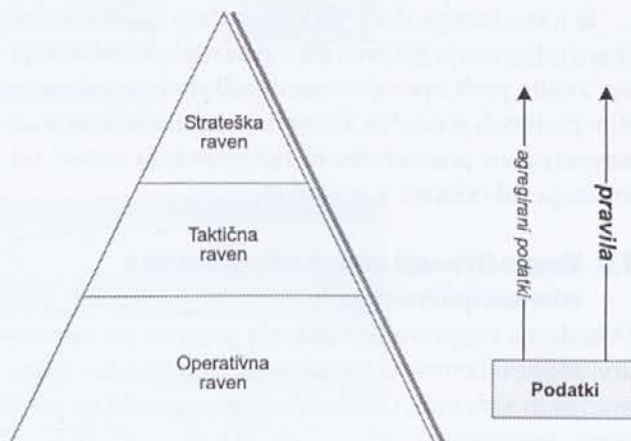
podatkih dejstvo, da se na podlagi podatkov za potrebe podpore odločanju ne pridobivajo več le agregirani podatki, temveč tudi modeli in pravila v okviru njih (slika 5).

4 SKLEP

Odkrivanje zakonitosti v podatkih je področje, ki se v okviru informacijskih sistemov vse bolj uporablja za podporo odločanju. Nekatere organizacije se za uporabo odkrivanja zakonitosti v podatkih odločijo za dosego konkurenčne prednosti pred konkurenco, druge so v to prisiljene za preživetje na trgu. V vsakem primeru pa predstavlja uporaba odkrivanja zakonitosti v podatkih višjo kakovostno raven podpore odločanju, saj omogočajo odkriti vzorci in pravila tako dodatno podlago za sprejemanje odločitev kot tudi akumuliranje znanja, kar omogoča kakovostnejše odločanje v prihodnosti.

Pojavitev vmesnikov in njihova standardizacija omogočata razvoj aplikativnih sistemov odkrivanja zakonitosti v podatkih, ki predstavljajo nov tip sistemov za podporo odločanju. Njihova prednost glede na interaktivna orodja je predvsem v tem, da omogočajo

izdelavo lastne baze pravil, komentiranje modelov in pravil ter pregledovanje pravil [Aggrawal 2002]. Poleg tega omogočajo delitev vlog na skrbnika modelov in analitika, kar omogoča porazdelitev znanj o odkrivanju zakonitosti v podatkih z večjo koncentracijo pri skrbniku modelov in manjšo pri analitiku. Prednost uporabe aplikativnih sistemov odkrivanja zakonitosti v



Slika 5: Prikaz pravil kot nove kategorije, ki predstavlja podlago za odločitvene procese

podatkih in prek njih pridobljenih pravil pa predstavlja tudi možnost uporabe pravil nad novimi podatki s strani drugih aplikativnih sistemov v okviru informacijskega sistema. Vmesniki in njihova standardizacija bodo vsekakor vplivali na razširitev uporabe odkrivanja zakonitosti v podatkih v okviru informacijskih sistemov.

5 LITERATURA

- [Academy] Agent Academy, IST Projekt (IST-2000-31050): A Data Mining Framework for Training Intelligent Agents, <http://agentacademy.iti.gr/>
- [Aggrawal 2002] AGRAWAL, C.C. (2002). Towards Effective and Interpretable Data Mining Visual Interaction, ACM SIGKDD Explorations Newsletter, 3(2), pp.11-22
- [Alter 2002] ALTER, S. (2002). Information Systems: The Foundation of e-Business, Addison-Wesley
- [Apte 2002] APTE, C., LIU, B., PEDNAULT, E.P.D. in SMYTH, P. (2002). Business Applications of Data Mining, Communications of the ACM, 45(8), pp.49-53
- [Bajec 2001a] BAJEC, M. (2001). Opredelitev izhodišč za celovito obnavljanje poslovnih pravil v organizacijah, Doktorska disertacija, Univerza v Ljubljani
- [Bajec 2001b] BAJEC, M. in KRISPER, M. (2001). Managing Business Rules in Enterprises, Elektrotehniški vestnik, 68(4), pp 236-241
- [Bayardo 2001] BAYARDO, R. in GEHRKE, J.E. (2001). Report on the Workshop on Research Issues in Data Mining and Knowledge Discovery Workshop (DMKD 2001), SIGKDD Explorations, 3, pp.43-44
- [Bohanec 2001] BOHANEK, M. (2001). What is Decision Support?, Proceedings of Information Society IS-2001: Data Mining and Decision Support in Action!
- [Crisp-dm] CHAPMAN, P., CLINTON, J., KERBER, R., KHABZA, T., REINARTZ, T., SHEARER, C. in RUDIGER, W. (2000). CRISP-DM 1.0, <http://www.crisp-dm.org>
- [Grossman 2002] GROSSMAN, R.L., HORNICK, M.F. in MEYER, G. (2002). Data Mining Standards Initiatives, Communications of the ACM, 45(8), pp.59-61
- [Hand 1999] HAND, D.J. (1999). Statistics and Data Mining: Intersecting Disciplines, ACM SIGKDD Explorations Newsletter, 1(1), pp.16-19
- [Heinrichs 2003] HEINRICH, J. in LIM, J.S. (2003). Integrating Web-based Data Mining Tools with Business Models for Knowledge Management, Decision Support Systems, 35, 103-112
- [Hirji 2001] HIRJI, K.K. (2001). Exploring Data Mining Implementation, Communications of the ACM, 44(8), pp.87-93
- [Holsheimer 1998] HOLSHEIMER, M. in SIEBELS, A. (1998). Data Mining – The Search for Knowledge in Databases, <http://www.cwi.nl>
- [Holsheimer 1999] HOLSHEIMER, M. (1999). Data Mining by Business Users: Integrating Data Mining in Business Process, Proceedings International Conference on Knowledge Discovery and Data Mining KDD-99, pp. 266-291
- [Hornick 2003] HORNICK, M.L. (2003). Java Data Mining (JSR-73): Overview and Status Abstract, <http://www.ncdm.uic.edu/workshops/dm-ssp03/hornick-abstract.htm>
- [Int01] Skripta za vaje za predmet Odločitveni sistemi, <http://lisa.uni-mb.si/student/predmeti/mosVS/vaje/gradiva/Podatkovno%20rudarjenje.ppt>
- [Int02] Oracle 9i Daily Feature: Oracle 9i Data Mining, <http://otn.oracle.com/products/oracle9i/daily/apr02.html>
- [Int03] Oracle Data Mining for Java (DM4J), <http://otn.oracle.com/products/bi/odm/9idm4jv2.html>
- [JSR-73] JSR-73 Expert Group (2002). Java Specification Request 73: Java Data Mining (JDM)
- [Kohavi 2002] KOHAVI, R., ROTHLEDER, N.J. in SIMOUNDIS, E. (2002). Emerging Trends in Business Analytics, Communications of the ACM, 45(8), pp.45-48
- [Kononenko 1997] KONONENKO, I. (1997). Strojno učenje, Založba FE in FRI.
- [Kononenko 2001] KONONENKO, I. (2001). Machine Learning for Medical Diagnosis: History, State of the Art and perspective, Artificial Intelligence in Medicine, 23(1), pp.89-109
- [Kukar 1997] KUKAR, M., GROŠELJ, C., KONONENKO, I. in FETTICH, J. (1997). An Application of Machine Learning in the Diagnosis of Ischaemic Disease, Proceedings of the 6th Conference on Artificial Intelligence in Medicine Europe, pp. 461-464
- [Kukar 2002] KUKAR, M. in KONONENKO, I. (2002). Reliable Classifications with Machine Learning, Proceedings of the 13th European Conference on Machine Learning, pp. 291-231
- [Laudon 1998] LAUDON, K.C. in LAUDON, J.P. (1998). Management Information Systems, Prentice Hall
- [Little 2002] LITTLE, B.B., JOHNSTON, W.L. in LOVELL A.C. (2002). Collusion in The U.S. Crop Insurance Program: Applied Data Mining, Proceedings International Conference on Knowledge Discovery and Data Mining KDD-2002, pp. 594-598
- [Moyle 2002] MOYLE, S., BOHANEK, M. in OSTROWSKI, E. (2002). Large and Tall Buildings: A Case Study in the Application of Decision Support and Data Mining, Proceedings of the Second International Workshop Integration and Collaboration Aspects of Data Mining, Decision Support and Meta-Learning, pp.88-99
- [Rupnik 2001] RUPNIK, R. (2001). Uvedba CRM – strateška prednost pred konkurenco, Zbornik konference INFOS 2001, pp.25
- [SolEuNet] Data Mining and Decision Support for Business Competitiveness: A European Virtual Enterprise, IST projekt (IST-1999-11495), <http://soleunet.ijs.si/website/html/euproject.html>
- [Slovar] Slovar informatike, Slovar jezikovne sekcije Slovenskega društva informatika, http://www.islovar.org/iskanje_enostavno.asp
- [Westphal 1998] WESTPHAL, C., BLAXTON, T. (1998). Data Mining Solutions, Wiley Computing Publishing

Rok Rupnik je zaposlen kot višji predavatelj na Fakulteti za računalništvo in informatiko v Ljubljani, kjer je leta 1998 magistriral in leta 2002 doktoriral. Njegovo raziskovalno področje obsega metodologije razvoja informacijskih sistemov, širše področje elektronskega poslovanja, inteligentne agente, odkrivanje zakonitosti v podatkih, strateško planiranje informacijskih sistemov ter mobilne aplikacije in mobilno poslovanje. V svoji karieri je sodeloval pri več razvojnih projektih informacijskih sistemov, na večini od njih v vlogi koordinatorja in vodje razvoja. Z vidno vlogo ali vlogo vodje oz. koordinatorja projekta pa je sodeloval tudi na različnih ostalih projektih širšega področja informacijskih sistemov: izdelave strateških planov razvoja informatike za večje poslovne sisteme, izdelave strateških študij, raziskovalnih projektih itn. Med drugim je tudi soavtor in urednik Enotne metodologije razvoja informacijskih sistemov. Je član Slovenskega društva INFORMATIKA, IEEE, ACM, združenja AIS (Association for Information Systems) in ustanovitveni član slovenske sekcije PMI (Project Management Institute).

Marjan Krisper je zaposlen kot docent na Fakulteti za računalništvo in informatiko v Ljubljani. Njegovo raziskovalno področje obsega metodologije razvoja informacijskih sistemov, področje elektronskega poslovanja, strateško planiranje informacijskih sistemov, agilne metodologije, poslovna pravila in področje prenove poslovnih procesov. V svoji karieri je bil vodja več razvojnih projektih informacijskih sistemov in projektov izdelave strateških planov informacijskih sistemov. Med drugim je tudi nosilec ideje, soavtor in urednik Enotne metodologije razvoja informacijskih sistemov. Je član Slovenskega društva INFORMATIKA, ACM, združenja AIS (Association for Information Systems) in ustanovitveni član slovenske sekcije PMI (Project Management Institute).