

Univerzita Pardubice
Fakulta ekonomicko-správní

Použití RM ve forenzním auditu

Bc. Denisa Kylarová

DIPLOMOVÁ PRÁCE
2009

Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky
Akademický rok: 2008/2009

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Denisa KYLAROVÁ**

Studijní program: **N6209 Systémové inženýrství a informatika**

Studijní obor: **Pojistné inženýrství**

Název tématu: **Použití RM ve forenzním auditu**

Z á s a d y p r o v y p r a c o v á n í :

Pojištění
Riziko
Pojistné podvody
Forenzní audit v pojišťovně
Indikátory pojistných podvodů
Rough množiny
Použití RM ve forenzním auditu

Rozsah grafických prací:

Rozsah pracovní zprávy:

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

Ducháčková, E., Principy pojištění a pojišťovnictví, Ekopress, Praha, 2005, ISBN 80-86119-92-0.

www.cap.cz

KOMOROWSKI, Jan, POLKOWSKI, Lech, SKOWRON, Andrzej. Rough Sets: A Tutorial. In: Rough-Fuzzy Hybridization : A New Trend in Decision-Making. Singapore : Springer-Verlag, [1998], s. 3-98.

POLKOWSKI, L. Rough Sets – Mathematical foundations. Heidelberg: Physica-Verlag company, 2002. ISBN 3-7908-1510-1.

ĀLEZAK, Dominik, et al. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing : Part II. Berlin : Springer-Verlag, 2005. 738 s. ISBN 3-540-28660-8.

SANKAR, K. Pal, SIMON, C. K. Shiu. Foundations of Soft Case-Based Reasoning. 2nd edition. [s.l.] : A Wiley-Interscience publication, c2004. ISBN 0-471-08635-5.

Vedoucí diplomové práce:

Ing. Pavel Jirava, Ph.D.

Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce:

6. října 2008

Termín odevzdání diplomové práce:

1. května 2009



doc. Ing. Renáta Myšková, Ph.D.

děkanka

L.S.



doc. Ing. Jiří Křupka, Ph.D.

vedoucí ústavu

V Pardubicích dne 6. října 2008

Prohlašuji:

Tuto práci jsem vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 21. 8. 2009

Denisa Kylarová

Ráda bych touto cestou poděkovala svému vedoucímu práce Ing. Pavlu Jiravovi, Ph. D. za jeho cenné rady, připomínky, trpělivost a vstřícnost. Dále bych chtěla poděkovat své rodině a přátelům za podporu v době studia.

Anotace

Tato diplomová práce se zabývá odhalováním pojistných podvodů pomocí RM. První kapitola se věnuje základním pojmům z pojištění a rozdělení na sociální a komerční pojištění. Ve druhé kapitole je charakterizován pojem riziko a jeho základní členění. V třetí části této práce se seznámíme s forezním auditem v pojišťovně. Teorií RM a jejími matematickými základy se seznámíme ve čtvrté kapitole. Poslední, a to zásadní kapitola, se zabývá použitím RM ve forezním auditu, jako modelu pro detekci pojistných podvodů.

Klíčová slova

riziko, forezní audit, pojistný podvod, RM

Title

Applying Rough Sets in Fraud Detection

Annotation

This thesis deals with rough sets used for insurance fraud detection. The first chapter examines the basic of insurance and its dividing in national insurance and commercial insurance. The second chapter is characterized by risk and its classifying. In the third chapter of this thesis we acquaint with forensic audit in insurance company. In the fourth chapter we acquaint with theory of rough sets and its mathematical foundations. The last, a major chapter deals with rough sets used for forensic audit, like model for fraud detection.

Keywords

risk, forensic audit, insurance fraud, rough sets

OBSAH

Obsah	6
Úvod.....	7
1 Pojištění.....	9
1.1 Základní pojmy	9
1.2 Členění pojištění	11
1.2.1 Sociální (veřejná) pojištění.....	12
1.2.2 Komerční (soukromé) pojištění.....	12
2 Riziko.....	14
2.1 Členění rizik.....	14
2.1.1 Riziko interní a externí.....	15
2.1.2 Riziko podle velikosti.....	15
2.1.3 Riziko pojistitelné a nepojistitelné	17
2.1.4 Riziko skutečné a spekulativní	17
2.2 Řízení rizika.....	18
2.2.1 Fáze risk managementu	18
3 Forenzní audit v pojišťovně	20
3.1 Audit	20
3.1.1 Interní audit	20
3.1.2 Forenzní audit.....	21
3.2 Pojistné podvody	21
3.2.1 Indikátory pojistných podvodů u pojištění motorových vozidel	23
3.2.2 Vyšetřování pojistných podvodů.....	25
4 RM	27
4.1 Matematické základy RM.....	28
4.1.1 Základní pojmy	28
4.1.2 Relace.....	29
4.1.3 Aproximace množiny	29
4.1.4 Informační systém.....	31
4.1.5 Příklad – RM	31
4.2 Aplikace a softwarové nástroje.....	34
4.2.1 Softwarové nástroje.....	35
5 Použití RM ve forenzním auditu	38
5.1 Příprava dat.....	39
5.1.1 Vlastní příprava dat	40
5.2 Modelování	49
5.2.1 Vlastní postup modelování	49
5.3 Výstupy z modelování	58
Závěr	62
Seznam literatury	64
Seznam obrázků	66
Seznam tabulek	66
Seznam použitých zkratk.....	67
Seznam příloh	68
Přílohy.....	69

ÚVOD

V současné době se počet pojistných podvodů oproti minulosti rapidně zvyšuje. Jejich vyšetřování je poměrně nákladnou záležitostí, a to z toho důvodu, že se prošetřují ručně bez použití softwarových produktů, kterých na trhu není dostatek. Z tohoto důvodu mě tato oblast zaujala, pro experiment, zda by se daly pojistné podvody odhalovat pomocí RM. Zavedení modelu pro detekci podvodů páchaných nejčastěji na pojištění motorových vozidel by mohlo pro pojišťovnu znamenat ušetření nejen času, ale i finančních nákladů.

Během roku 2008 pojišťovny vyšetřovaly téměř 5 tisíc případů, které vykazovaly známky spáchání pojistného podvodu. Oproti roku 2007 se celkový počet šetřených případů zvýšil o 7 % [1].

V porovnání s předchozím rokem 2007 mírně poklesl počet vyšetřovaných událostí především u pojištění majetku. Na rozdíl od zvýšení pokusů o pojistný podvod v oblasti pojištění vozidel. S těmito faktory souvisí také to, že v loňském roce nedošlo k žádným významnějším živelným pohromám. Ty zpravidla stojí za vysokými škodami v oblasti pojištění majetku, kde se v jejich rámci běžně zvyšuje riziko pokusů o pojistný podvod [1].

V počtu prozkoumání případů, na kterých bylo pácháno nejvíce pojistných podvodů, stále vede pojištění motorových vozidel. V loňském roce prošetřily pojišťovny v rámci této oblasti pojištění více než 3,5 tisíce šetření. Na základě těchto šetření uchránily před neoprávněným vyplacením téměř 350 milionů korun [1].

Pojišťovnám se také podařilo, ve spolupráci s Policií ČR, úspěšně odhalit několik organizovaných skupin pachatelů. Tyto skupiny se nejvíce specializovaly na páchání pojistného podvodu v oblasti pojištění vozidel [1].

V poslední době dochází k pojistným podvodům i v segmentu životního pojištění, kde často dochází k velmi kuriózním případům sebepoškozování [1].

Pro příklad, jenom samotná Česká pojišťovna, a. s. dokázala v roce 2008 odhalit celkem 703 případů pojistných podvodů. Zvládla tak uchránit na pojistném plnění 274 milionů korun. Za rok 2007 se jednalo o 739 případů za 264 milionů korun. Většinou se jednalo o nadhodnocené nebo falešné pojistné události, nejvíce se jich týkalo oblasti motorových vozidel a pojištění majetku [1].

Pojistný podvod je však trestným činem, za který hrozí pachatelům dvouletá sazba odnětí svobody. Osm let může podvodník strávit ve vězení, pokud škoda dosáhne výše půl milionu korun. A v případě spáchání pojistného podvodu za pět a více milionů korun, může být sazba až dvanáct let [1].

Česká asociace pojišťoven se dlouhodobě snaží o maximální využití informačních technologií v boji proti pachatelům trestného činu podvodu. Taktéž podporuje zavádění nových softwarových produktů, které mají pomoci pojišťovnám při detekci podvodu [1].

Vzhledem k tomu, že v současnosti, je na trhu jen velmi málo softwarových nástrojů, které mají za úkol odhalovat pojistné podvody, je cílem této práce sestavení modelu pro detekci těchto podvodů, sestaveného dle základu RM.

Dále česká asociace pojišťoven usiluje o zlepšení legislativy, která by neměla otevírat prostor pro působení nových podvodníků. Naopak by se mělo poskytnout více zákonných prostředků pojišťovnám, aby byly schopny s rizikem podvodu bojovat a eliminovat jej [1].

1 POJIŠTĚNÍ

Prvotní myšlenka, filosofie pojištění měla humánní záměr. Hlavním cílem pojištění byla pomoc v nouzi, vzájemná výpomoc v situacích, kdy se jednotlivci nemohou ubránit vzniku škod a sami krýt jejich následky [2].

Pojištění představuje finanční službu. Za úplatu je nám poskytnuta pojistná ochrana. Přesouváme rizika na instituce, která provozují pojištění. Stručně můžeme pojištění charakterizovat jako nástroj finanční eliminace nežádoucích důsledků nahodilých jevů. Pomocí pojištění nemůžeme odvrátit výskyt náhodných událostí, ale můžeme finančně nahradit ztráty, které vznikly při dopadu nahodilých událostí [2], [3].

V tržní ekonomice pojišťovací činnost stabilizuje ekonomickou úroveň podniků a životní úroveň obyvatelstva v případě neočekávané události. Pojišťovnictví je podporováno i ze strany státu, prostřednictvím daní a zavedením zákonných a smluvních povinných forem pojištění. Pojištění je v zásadě založeno na principu solidarity, podmíněné návratnosti a zároveň neadekvátní návratnosti vložených prostředků [2].

1.1 Základní pojmy

Pojistník

Je osoba, která s pojistitelem uzavřela pojistnou smlouvu. Pojistníkem může být osoba fyzická i osoba právnická. Za poskytnutou pojistnou ochranu má povinnost platit pojistné a je podepsán na pojistné smlouvě [4], [5].

Pojištěný

Pojištěný je osoba, na jejíž zdraví, život, odpovědnost za škody, majetek či jiné hodnoty pojistného zájmu, se pojištění vztahuje. Tato osoba má nárok na pojistnou ochranu, a to i když pojištění nesjednala sama, ale pomocí jiné osoby – pojistníkem [4], [5].

Pojistitel

Pojistitelem je právnická osoba, ta má oprávnění provozovat pojišťovací činnost dle zákona č. 363/1999 Sb., o pojišťovnictví a o změně některých souvisejících zákonů (zákon

o pojišťovnictví), ve znění pozdějších předpisů. Dle tohoto zákona je pojistitelem pojišťovna, nebo jiná instituce, které bylo uděleno povolení k provozování pojištění [4], [5].

Obmyšlený

V případě pojištění, ve kterých je kryto riziko úmrtí (životní pojištění), vzniká nárok na pojistné plnění, které získá osoba určená pojistníkem v pojistné smlouvě. Tato osoba se nazývá obmyšlený [4], [5].

Oprávněná osoba

Oprávněná osoba je ta osoba, které případně v případě vyskytnutí pojistné události nárok na pojistné plnění [4], [5].

Pojistná smlouva

Je to dokument, na jehož základě vzniká smluvní pojištění fyzických nebo právnických osob. Jedná se o dvoustranný právní úkon, který musí být vyhotoven v písemné formě podle platných právních předpisů [4], [5].

Pojistná smlouva představuje smlouvu o službě, ve které se pojišťovna zavazuje při vzniku nahodilé události umožnit ve sjednaném rozsahu výplatu plnění, a na druhé straně, pojistník je za tuto službu povinen platit pojistné. Její součástí jsou pojistné podmínky, jak všeobecné, tak i zvláštní pojistné podmínky [4], [5].

Pojistná smlouva musí vždy obsahovat [4], [5]:

- určení pojistníka, pojistitele a oprávněné osoby,
- určení, o jaké pojištění se jedná, zda škodové, či obnosové,
- vymezení pojistné události a pojistného nebezpečí,
- celkovou výši pojistného, jeho splatnost, a údaj o tom, zda je to jednorázové nebo běžné pojistné,
- vymezení pojistné doby a délka trvání pojistné smlouvy,
- při pojištění osob, v případě podílení oprávněné osoby na výnosech, i způsob podílení se na nich,
- pojistné podmínky (připojené k pojistné smlouvě).

Pojistná doba

Jedná se o celkovou délku, na kterou bylo pojištění sjednáno. Je uvedena v pojistné smlouvě. Pojištění může být sjednáno na dobu určitou a neurčitou. Pojistná doba se dále člení na jednotlivá pojistná období [4], [5].

Pojistný kmen

Pojistný kmen představuje portfolio spravovaných pojištění. Toto portfolio tvoří souhrn pojistných smluv spravovaných podle jednotlivých druhů pojištění [4], [5].

Pojistné riziko

Souhrn rizik, která jsou kryta příslušným druhem pojištění. Je to míra pravděpodobnosti vzniku pojistné události vyvolané pojistným nebezpečím [4], [5].

Pojistné podmínky

Podmínky, které zpracovává pojistitel pro uzavírání pojistných smluv pro jednotlivá pojistná odvětví, pro skupiny těchto odvětví nebo pro jednotlivé typy pojištění uzavíraných v rámci pojistného odvětví. V pojistných podmínkách je obsažena právní úprava určitého pojistného produktu. Používají se všeobecné podmínky i zvláštní pojistné podmínky [4], [5].

1.2 Členění pojištění

V současné době se, v České republice, pojištění na veřejné a soukromé. Veřejné pojištění je známé také jako sociální pojištění. To zahrnuje úhradu sociálních rizik v rozsahu, který je dán státem. Jedná se o pojištění povinné. Soukromé, neboli komerční pojištění, obsahuje zabezpečení rizik ekonomickým subjektům (jak fyzickým tak i právnickým osobám) dle jejich vlastních rozhodnutí a potřeb (kromě některých povinných pojištění) [3], [6].

Z právního hlediska dělíme pojištění na

- dobrovolná (závisí pouze na rozhodnutí pojistníka, sjednává se pojistná smlouva),
- povinná:
 - o povinná smluvní (povinnost sjednat pojistnou smlouvu pro dané subjekty),
 - o zákonná (vyplývá ze zákona a pojistná smlouva se nesjednává) [3].

1.2.1 Sociální (veřejná) pojištění

V rámci tohoto pojištění jsou kryta rizika, která mají charakter sociální, a z nich vyplývající potřeby. Jsou to zejména tyto potřeby [3]:

- v souvislosti s dlouhodobou pracovní neschopností,
- v souvislosti s krátkodobou pracovní neschopností,
- v souvislosti s pracovními úrazy a nemocemi z povolání,
- zdravotní péče,
- v souvislosti s nezaměstnaností.

1.2.2 Komerční (soukromé) pojištění

Toto pojištění v sobě zahrnuje řadu produktů. Typickým rysem pro komerční pojištění je smluvní podoba pojištění. Soukromé pojištění je časté jako dobrovolné, ale může být i povinné ze zákona [3].

Blíže si představíme dva produkty komerčního pojištění, havarijní pojištění a tzv. povinné ručení.

Pojištění motorových vozidel (havarijní pojištění)

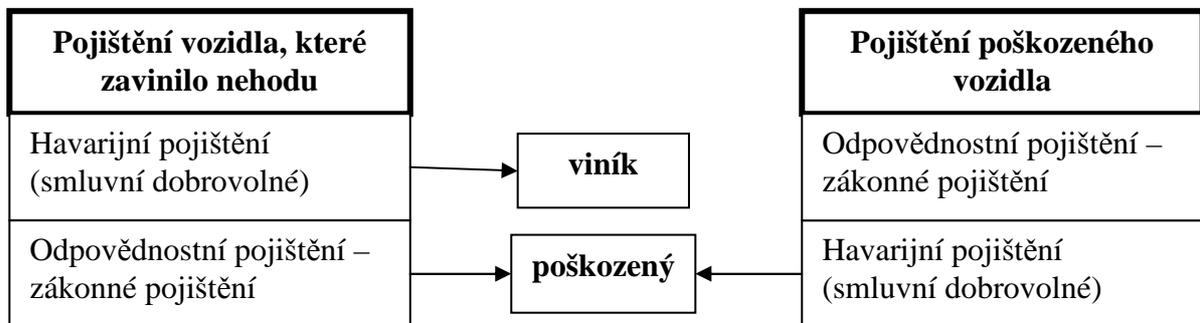
Poskytuje pojistné plnění majiteli nebo držiteli vozidla v případě pojistné události, kterou může být poškození, zničení nebo odcizení motorového vozidla. Škody mohou vzniknout v důsledku [7]:

- jakékoliv živelné události,
- krádeže, loupeže, vandalství, poškození, zničení nebo odcizení motorového vozidla, neoprávněné používání vozidla,
- dopravní nehodou, způsobenou jakoukoliv událostí, včetně nesprávného jednání řidiče motorového vozidla.

Pojištění odpovědnosti za škodu způsobenou provozem vozidla (povinné ručení)

Jedná se o povinné smluvní pojištění. Povinné ručení kryje v celé České republice za pojištěného všechny škody, které pojištěný způsobil, třetím osobám, provozem vozidla, kterého je vlastníkem [7].

Vztah smluvního povinného pojištění odpovědnosti za škodu způsobenou provozem vozidla (povinné ručení) a smluvního dobrovolného pojištění motorových vozidel (havarijní pojištění) je znázorněn na obrázku 1 Obrázek 1.



Obrázek 1: Vztah povinného ručení a havarijního pojištění [7]

2 RIZIKO

Původně z italského slova „risico“, je možné volně přeložit jako nebezpečí nezdaru, škody, ztráty, dále také jako možnost škody, ztráty, zranění, nezdaru, hazard či možnost, že s určitou pravděpodobností vznikne odlišná událost [2], [3], [7].

V arabštině („risk“), toto slovo, původně označovalo nejen nepříznivou, ale i příznivou událost v životě člověka. S postupem času se pojem riziko začalo používat v souvislosti s nepříznivou událostí [7].

Nejdříve bylo užívání slova riziko spojené hlavně s pojišťovnictvím. Teprve s rozvojem společnosti se tento pojem rozšířil i do jiných vědních disciplín [2].

Riziko je možné chápat jako pravděpodobnost, že dojde ke škodlivé události. Jedná se o reakci na hrozbu¹. Mimo pojišťovnictví můžeme riziko považovat za nebezpečí, např. že výsledek podnikání nebo činnosti nebude ve skutečnosti takový, jak se předpokládalo v době rozhodnutí. Všeobecně lze riziko definovat jako určitou objektivní pravděpodobnost, že vznikne událost, která se liší od toho, co si přejeme. Riziko pro nás představuje nejistotu, kterou můžeme měřit, pomocí určité hodnoty pravděpodobnosti [2], [3], [7].

Riziko R může být vyjádřeno mnoha faktory, za základní považujeme velikost škody Z , pravděpodobnost vzniku škody p a čas t . Formálně můžeme riziko vyjádřit funkcí

$$R = f(Z, p, t, \text{další faktory}). \quad (1)$$

Pojistné riziko je míra pravděpodobnosti, že vznikne pojistná událost [3].

2.1 Členění rizik

V naší ekonomice se nachází více druhů rizik. Tato rizika můžeme začlenit do následujících skupin [2]:

- obchodně finanční,
- sociálně – politická,
- živelná,
- ostatní.

Rizika můžeme také dělit podle jiných kritérií. Zatřídít je můžeme např. z pohledu pojišťovny nebo také z pohledu podniků, a to následovně jako riziko [7]:

- interní a externí,
- ovlivnitelné a neovlivnitelné,
- podle velikosti,
- pojistitelné a nepojistitelné,
- skutečné nebo spekulativní,
- systematické a nesystematické,
- subjektivní a objektivní,
- fyzické a morální,
- elementární a specifické,
- atd.

2.1.1 Riziko interní a externí

To, jak podnik dokáže nakládat s riziky a eliminovat je, do jisté míry také závisí, jak na něj působí. Zda se jedná o vnitřní (interní) nebo vnější (externí) rizika. To jsou také faktory, které mají vliv na podnik [7].

Interní riziko

Sem patří ta rizika, která se projevují uvnitř podniku. Tato rizika je vedení podniku schopno řídit a ovlivňovat. Příkladem mohou být zaměstnanci, výrobní procesy, finanční síla organizace, apod. [7].

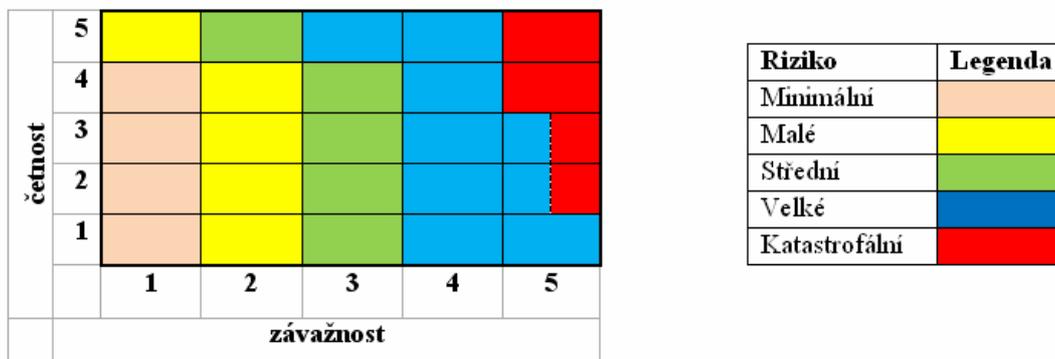
Externí riziko

Jedná se o faktory, které na podnik působí zvenčí. Tyto vlivy není podnik schopen sám řídit a ovládat. Patří sem např. obchodní podmínky, ochrana životního prostředí, bezpečnostní situace, atd. [7].

2.1.2 Riziko podle velikosti

Rizika je možné třídit na základě závažnosti a četnosti, kde závažnost bývá vystižena velikostí nákladů na škodní událost aj. Pro rozpoznání rizik, která mohou nejvíce a nejčastěji ohrozit

podnik či občana, nebo která budou nejdražší je důležitá klasifikace těchto rizik podle velikosti. Na následujícím obrázku můžeme vidět členění rizik podle jejich velikosti, viz obrázek 2 [7].



Obrázek 2: Členění rizik podle velikosti [7]

Minimální riziko

Pro nás znamená zanedbatelný náklad na zvládnutí vzniklé škody. Kompenzace škod nám neovlivní náklady, pokud by je ovlivnila, tak pouze v malém rozsahu. Není potřeba čerpat z finančních rezerv. Velikost škody nemá vliv na proces ve výrobě [7].

Malé riziko

Tvoří velmi nízký náklad na překonání vzniklých škod. Pro vyrovnání škody čerpáme z finanční rezervy jen v malém rozsahu. Vliv na výrobní proces je jen velmi malý [7].

Střední riziko

Už může značit patrné náklady, které jsou potřebné na odstranění škody. Tyto náklady mohou výrazně ovlivnit životní úroveň podniku či občana a je potřebné sáhnout na finanční rezervy. Rozsah škody už může zpomalit výrobní proces [7].

Velké riziko

Znázorňuje vysoké poškození hmotných a jiných hodnot podniku nebo občana. Vyrovnání škod probíhá pomocí velké části finančních a hmotných rezerv, případně jiných forem financování škod (pojištění, samopojištění). Velikost škody může dlouhodobě ohrozit hospodářský proces podniku [7].

Katastrofální riziko

Katastrofální riziko značí mimořádně rozsáhlé poškození hmotných a jiných hodnot podniku nebo majetku občana. Směřuje k dlouhodobému zastavení výroby nebo ke zřícení celého systému. Likvidace škod je možná jen s využitím všech hmotných a finančních rezerv a s jinými formami financování jako je pojištění [7].

2.1.3 Riziko pojistitelné a nepojistitelné

Toto členění používají především pojišťovny při pojišťování a při uvažování o možné pojistné ochraně. Můžeme uvažovat nad tím, že neexistuje riziko, které by nebylo pojistitelné, a jeho pojistitelnost je ovlivněna jedinec cenou pojišťovací služby. V praxi však dochází k tomu, že pojišťovna nepojistí jakékoli riziko a samotné poskytnutí pojistné ochrany musí důkladně zvažovat [7].

Pojistitelné riziko musí splňovat tato kritéria:

- riziko musí být identifikovatelné,
- ztráta musí být vyčíslitelná,
- riziko musí být ekonomicky přijatelné pro pojišťovnu,
- projev rizika musí být nahodilý.

Nepojistitelné riziko je takové riziko, které tyto podmínky nesplňuje. Není zde možnost ho krýt pomocí komerčního pojištění [3], [7].

2.1.4 Riziko skutečné a spekulativní

Základní rozdíl mezi skutečným a spekulativním rizikem je, že skutečná rizika jsou obvykle pojistitelná, zatímco spekulativní rizika nikoli. Můžeme říci, že není možné pojistit ta rizika, jejichž výsledkem může být zisk. Spekulativní rizika podstupuje člověk dobrovolně, jelikož doufá, že dosáhne zisku [3], [7].

Riziko skutečné

Při realizaci rizika vznikne výhradně negativní odchylka od cíle. Riziko skutečné můžeme jinak nazývat také riziko čisté. Příkladem může být riziko havárie motorového vozidla, požár, úraz v zaměstnání. Toto všechno jsou skutečná rizika, při kterých nám hrozí nebezpečí ztrát. Neexistuje zde prvek zisku, buď se stane nehoda (požár, úraz) nebo ne. Když k dané události nedojde, stav se nezmění [3], [7].

Riziko spekulativní

Hovoříme o záměrném riziku, kde existuje možnost zisku. S realizací rizika nám mohou vzniknout jak záporné tak i kladné odchylky od cíle. Jedná se o situace spojené s hraním hazardních her, sázením, spekulacemi na burze, apod. Příslušný subjekt riziko dobrovolně podstupuje [3], [7].

2.2 Řízení rizika

V dnešní době, je snaha zvládat rizika pomocí vědeckých přístupů. Tento postoj vedl ke vzniku speciálního oboru pro řízení rizik – risk management (rizikový management) [2], [3], [7].

Obor řízení rizika se snaží o racionální jednání v rizikových situacích tak, aby byla chráněna současná a budoucí aktiva podniku [2], [3], [7].

Risk management zastupuje cílevědomé aktivity od předcházení vzniku a realizace rizik až po omezování rozsahu škod, ke kterým může dojít. Má se snažit o odhalení a zmírnění, pokud je to možné, všech nebezpečí hrozcí podnikatelské činnosti [2], [3], [7].

K úkolům manažera rizik patří především:

- maximální možné omezení pravděpodobnosti výskytu rizik,
- předvídání a organizování důsledků realizace rizik takovým způsobem, aby dopad (finanční, obchodní, lidský) byl pro podnik co nejnižší.

Cílem rizikového managementu je dosažení bezpečné činnosti při co nejnižších nákladech na zajištění této bezpečnosti [3].

2.2.1 Fáze risk managementu

Rizikový management představuje řadu činností, které může rozdělit do určitých fází. Základní fáze můžeme členit následovně [3]:

- identifikace rizik,
- ocenění a kvantifikace rizik,
- kontrola a financování rizik.

Identifikace rizika

V této fázi rozpoznáváme všechny rizikové faktory podniku. Zjišťujeme, která rizika mohou ohrožovat ekonomickou stabilitu [3].

Ocenění a kvantifikace rizik

Zjišťujeme, jakou váhu jednotlivá rizika mají, a jaký dopad může mít realizace rizik na finanční situaci daného podniku. Zpravidla se počítá s maximální možnou škodou, která může být realizací rizik způsobena [3].

Kontrola a financování rizik

Přijímáme opatření, kterými předcházíme škodám, a rozhodujeme o finanční kompenzaci důsledků negativních nahodilých událostí [3].

Předcházet rizikům lze pomocí [3]:

- strategických opatření (smluvní vyloučení odpovědnosti, používání bezpečnějších technologií, atd.),
- fyzických opatření (použití ochranných pomůcek, protipožárního zařízení, atd.).

Pokud nelze rizikům předejít pomocí strategických nebo fyzických opatření, má podnik možnost krýt rizika pomocí [3]:

- vlastních zdrojů (z běžných příjmů, samopojištění),
- přenesení rizik na jinou instituci, která poskytuje pojistnou ochranu (využití pojišťoven).

3 FORENZNÍ AUDIT V POJIŠŤOVNĚ

V této kapitole, si blíže představíme, co je to audit a jaký je rozdíl mezi interním a forezním auditem. Dále si v této části práce přiblížíme pojistné podvody a jejich vyšetřování.

3.1 Audit

Původně z latinského slova *auditus*, které znamená slyšení (auditoři informovali ve starém Římě senát o výši státního majetku). V dnešní době má audit význam jako úřední přezkoumání a posouzení dokumentů, zejména účtů, nezávislou osobou [8].

„Audit je v podstatě kritická analýza, která umožňuje ověřovat informace dodané podnikem a hodnotit operativní činnosti a systémy při jejich předávání.“ Lawrenc B. Sawyer

Hlavním účelem auditu je vyšetření, zda doklady podávají platné a spolehlivé informace o skutečnosti. Dalším úkolem také bývá ohodnocení, jak je kvalitní vnitřní kontrola firmy [8], [9].

Vedle finančního auditu, který se zabývá zejména účetními výkazy, může být audit zaměřen i na různé oblasti lidské činnosti. Mezi nejznámější druhy auditu patří: audit kvality, informační audit, ekologický audit, počítačový audit, atd. [8], [9].

Audit můžeme členit také podle toho, kdo ho provádí, viz tabulka 1. Tzn. na vnitřní, neboli interní audit, který provádějí zaměstnanci firmy. Nebo na vnější, nazývaný také externí audit, který provádí specializovaná firma [8], [9].

Tabulka 1: Rozdíl mezi interním a externím auditem [vlastní [8]]

Audit	Vnitřní (interní)	Vnější (externí)
Kdo ho provádí?	zaměstnanci firmy	externí akreditovaná firma
Kdo využívá výsledky?	pouze auditovaná organizace	jiné organizace (např. odběratelé)

3.1.1 Interní audit

Neměl by se zabývat prostou kontrolou. Jeho cílem je sloužit lépe vedení podniku jako nástroj pro dosažení efektivnosti řízení podniku. V rámci organizační struktury se nachází vedle nejvyššího řídicího stupně. Toto postavení by mělo internímu auditu zajistit nezávislost [9].

Interní audit je objektivní, ujišťující, nezávislá a konzultační činnost, která se zaměřuje na přidanou hodnotu a zlepšení provozu organizace. V pojišťovně IA nabízí managementu a vlastníkům pojišťovny přiměřené ujištění o stavu procesů a účinnosti kontrolních mechanismů [9].

Pomáhá pojišťovně k dosažení jejich cílů tím, že zavádí systematický a metodický přístup k hodnocení a zlepšení efektivnosti řízení rizik, řídicích a kontrolních procesů a správy a řízení organizace [9].

IA nám nenahrazuje vnitřní kontrolu. Interní auditor je kvalifikovaný pracovník, který předkládá managementu svůj názor. IA může formou konzultace, doporučení či pomocí auditorské zprávy poskytnout informace pro pochopení a zhodnocení rizika [9].

IA v pojišťovně se považuje za jeden z nejvýznamnějších a nejučinnějších nástrojů pro předcházení kriminálnímu jednání a odhalování pojistných podvodů. Není to však jeho primární cíl [9].

3.1.2 Forezní audit

Forezní je z latinského slova forensis, které znamená soudní. Vzniklo ze slova forum, což znamenalo veřejné prostranství, kde se konaly soudy. Obvykle označuje postupy a vědy, které souvisejí s vyšetřováním a soudním dokazováním, zejména v trestních záležitostech. Tyto vědy se souhrnně označují jako forenzika [10].

Pojišťovny vytvářejí specifické organizační jednotky, které bojují s kriminální činností klientů a zaměstnanců. Tato jednotka se nazývá forezní audit.

Interní audit se zabývá hlavně prevencí kriminální činnosti, naproti tomu se forezní audit zabývá především odhalováním kriminální činnosti. V pojišťovnictví to jsou pak zejména pojistné podvody.

3.2 Pojistné podvody

Podvod je trestný čin, který patří do kategorie majetkové kriminality. Podvody považují kriminalisté za velmi nebezpečnou formu trestné činnosti hlavně z důvodu jejich kvalifikovanosti a výskytu jak v oblasti obecné kriminalistiky, tak také v oblasti kriminality

„bílých límečků“. Takto se nazývá dle metodiky vyšetřování podvodů oblast hospodářské kriminality [11].

Podle § 250a se u základních skutkových podstat nevyžaduje, aby pojistným podvodem byla způsobena patrná škoda. Právě tato okolnost nebo alespoň úmysl škodu způsobit, je považován za nedostatek právní úpravy [11].

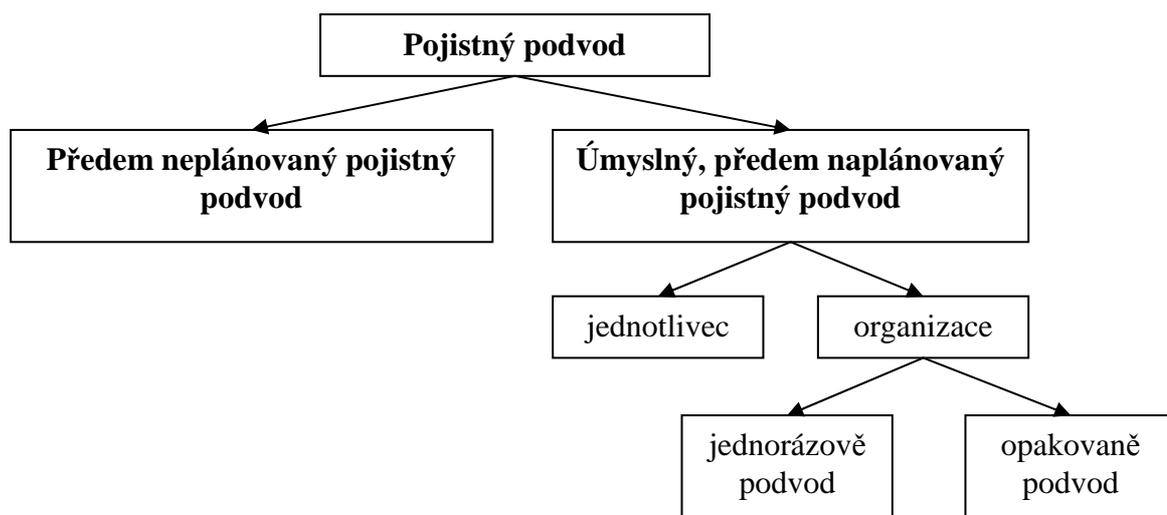
Dle znění § 250a jsou objektivní stránkou pojistného podvodu [11]:

- uvedení nepravdivých nebo hrubě zkreslených nebo zamlčení podstatných údajů v pojistné smlouvě či při uplatnění nároku na plnění,
- úmyslné vyvolání pojistné události, nebo udržování stavu vyvolaného pojistnou událostí v úmyslu zvýšit vzniklou škodu.

Z hlediska subjektivní stránky je pojistný podvod vždy úmyslným trestným činem. Předmětem pojistného podvodu je vždy pojistitel [11].

Pachatelem pojistného podvodu může být kdokoliv z účastníků pojistné smlouvy (pojištěný, pojistník, nebo osoba, která úmyslně vyvolá pojistnou událost, nebo úmyslně udržuje stav již vyvolaný, ať po dohodě s pojištěným, anebo bez takové dohody) [11].

Pojistné podvody můžeme rozčlenit do dvou základních skupin. Na předem neplánovaný pojistný podvod, kdy dotyčná osoba ani netušila, že provádí pojistný podvod a na úmyslný podvod, který byl předem naplánovaný. Toto rozdělení můžeme vidět na obrázku 3 [11].



Obrázek 3: Pojistný podvod [11]

Dalším členěním páchaní pojistných podvodů může být dle typu pojištění, tzn. na životní pojištění – pojištění osob nebo na neživotní pojištění [11].

Nejfrekventovanějším způsobem páchaní pojistného podvodu je u havarijního pojištění motorových vozidel a u pojištění odpovědnosti za škodu způsobenou provozem motorového vozidla [11].

3.2.1 Indikátory pojistných podvodů u pojištění motorových vozidel

Jak bylo uvedeno výše, nejčastěji páchanými pojistnými podvody jsou v praxi podvody související s pojištěním motorových vozidel [11].

Tyto podvody se páchají jednak v souvislosti s fingovanými krádežemi motorových vozidel, ale zároveň i s fingovanými dopravními nehodami, kterých v poslední době přibývá [11].

Podněcovatelem těchto pojistných podvodů bývá většinou sám pojištěný. Např. tím, že uvádí nepravdivé údaje při uzavírání pojistné smlouvy a při oznámení a následné likvidaci pojistné události. Využívá zfalšovaných dokumentů, úmyslně podává nepravdivá oznámení, předstírá pojistné události, maskuje datum vzniku pojistné události [11].

Indikátory jsou faktory, které mohou signalizovat, že pojistná událost má náznak podvodu. Mezi typické indikátory vztahující se ke škodným událostem, které vznikly v rámci pojištění motorových vozidel patří zejména tyto [11]:

Indikátory vzniku a průběhu pojištění

- absence pojistné historie z minulých pojištění,
- časté změny pojistitelů,
- několik škod během krátkého období,
- přerušení pojištění,
- neplacení nebo pozdní placení pojistného,
- řidičský průkaz je padělaný, aj.

Indikátory z hlediska místa, doby a příčiny nehody

- nehoda v noční době, rozporné místo nehody,
- nehoda v nefrekventovaném místě, beze svědků,
- střet najetím zezadu,

- vytačení ze silnice nebo oslnění vozidlem pojištěného bez kontaktu vozidel (nepřímé střety), střety se zvěří,
- chybí stopy nebo jich je nápadně mnoho, aj.

Indikátory z hlediska zúčastněných vozidel

- chybí STK,
- vozidlo škůdce je velmi ojeté,
- najetý počet km u poškozeného vozidla je v rozporu s jeho stářím či stavem,
- neúplná či sporná dokumentace k vozidlům, aj.

Indikátory z hlediska zúčastněných osob

- viník a poškozený jsou známí, příbuzní nebo mají jiné vazby,
- nápadný nesoulad mezi charakterem vozidla poškozeného a jeho sociální a ekonomickou situací,
- řidič není totožný s majitelem vozidla, aj.

Indikátory chování účastníků po nehodě a při likvidaci škody

- nehoda není hlášena policii, ačkoli měla být,
- zdržovací taktika vůči pojistiteli,
- oddalování termínu prohlídky vozidla po nehodě,
- neúměrný tlak na vyplacení pojistného plnění (zálohy),
- pozdní hlášení škodné události, aj.

Indikátory při dokumentaci a výpočtu škody

- sporná autentičnost dokladů,
- předložení pouze fotokopií dokladů,
- nadhodnocené nebo zfalšované faktury za opravu,
- výplata pojistného plnění na tři a více peněžních účtů, aj.

Pracovníci pojišťoven, kteří provádějí vyšetřování sporných pojistných událostí, zjistí pomocí indikátorů okolnosti svědčící o spáchání pojistného podvodu, nemohou vést vyšetřování trestných činů, protože nemají oprávnění policie [11].

Pojišťovna však pomocí svých vyšetřovatelů v zaměstnaneckém poměru a smluvních detektivů (v pojišťovně se nazývají forenzní detektivové) získává důkazy o pojistném podvodu. Pouze tito profesionálové, kteří většinou získali praxi v rámci své předchozí služby u policie, jsou schopni postupovat tak, aby nebyl porušen zákon [11].

Takto získané důkazy jsou pak předávány policii k dalšímu šetření, nebo mohou být použity při podání trestního oznámení státnímu zástupci [11].

3.2.2 Vyšetřování pojistných podvodů

V boji s pojistnými podvody spojují pracovníci pojišťoven své zkušenosti se zkušenostmi pracovníků policie. V rámci této činnosti se pojišťovny snaží zavádět takové pracovní postupy, které směřují k odhalení pojistného podvodu. Následně pak rozhodnou o tom, že uplatněný nárok na pojistné plnění není oprávněný [11].

Nejpodstatnější částí jsou zde indikátory, které pracovníky likvidační služby upozorňují na pojistné podvody [11].

Zvláštnosti předmětu vyšetřování

Při vyšetřování pojistného podvodu jako trestného činu, se nesetkáme se situací, kdy nám pachatel není znám. Zde však dochází k jiné situaci, a to že nemůže s určitostí tvrdit, že pojistný podvod byl spáchán [11].

Poškozeným je vždy pojistitel. Nemusí se ani zjišťovat motiv činu a podobně [11].

Podněty k vyšetřování

Odhalováním této trestné činnosti se zabývá především specializované pracoviště v pojišťovně (forenzní audit), dále pak kriminální policie [11].

Podnětem k začátku šetření trestného činu pojistného podvodu je oznámení poškozeného – pojistitele, že došlo ke spáchání trestné činnosti pojistného podvodu [11].

Po podání oznámení pojistitelem je však zpravidla věc podezření z pojistného podvodu předána k došetření pracovníkům kriminální policie [11].

Nejvyužívanější a nejrozšířenější opatření v boji proti pojistnému podvodu

Z pohledu pojišťoven to jsou zejména [11]:

- používání aktivních kontrolních opatření při uzavírání pojistných smluv a při likvidaci pojistných událostí,
- využívání specializovaných vyšetřovatelů uvnitř pojišťovny,
- využívání postupů externích subjektů.

Z pohledu trhu pak [11]:

- výměna informací a zkušeností mezi pojišťovnami,
- vytvoření centralizované informační kartotéky,
- navázání spolupráce s ostatními sektory finančních služeb,
- podporování informačních kampaní za účelem vyvolání zájmu nejširší veřejnosti o tuto problematiku (brožury, tisková prohlášení, atd.)

4 RM

Myšlenka RM byla navržena Zdislawem Pawlakem v roce 1982. Teorie RM patří mezi matematické metody, které pracují s neúplnými informacemi. Pojetí RM lze také nazvat modelem „přibližného usuzování“. RM jsou přiblížením ostrých množin. RM poskytují dolní a horní aproximaci původním množinám [12], [13], [14].

Teorie RM je přístupem k neurčitosti. Můžeme si to představit jako určitou realizaci této neurčitosti. Tato teorie je převážně používána pro odhalování nedokonalých znalostí z dat. Tento přístup má hodně předností, mezi nejvýznamnější klady patří [13], [14]:

- snadné porozumění,
- poskytuje vhodné algoritmy pro odkrytí skrytých charakteristik v datech,
- prokazuje vztahy, které nelze nalézt pomocí statistických metod,
- připouští použití jak kvalitativních, tak kvantitativních dat,
- nachází minimální množinu dat vhodnou pro klasifikaci (redukce dat),
- hodnotí důležitost dat,
- vytváří množiny z dat pomocí rozhodovacích pravidel.

Představa RM byla základem pro vývoj v teoretických výzkumech, jako jsou [12]:

- logika,
- algebra a topologie a aplikace výzkumu,
- objevování znalostí,
- data mining,
- teorie rozhodování,
- umělá inteligence.

RM se využívají v několika reálných aplikacích a v různých oblastech např. [12]:

- medicína,
- ekonomie, finančnictví a podnikání,
- případy týkající se životního prostředí,
- inženýrství,

- informační vědy,
- rozhodovací analýzy,
- sociální vědy a ostatní,
- molekulární biologie,
- chemie,
- farmacie,
- kultura a umění.

4.1 Matematické základy RM

V této kapitole si nejdříve přiblížíme základní pojmy.

4.1.1 Základní pojmy

Tato část je zpracovaná pomocí. V této kapitole si osvětlíme základní důležité pojmy, které potřebujeme znát pro přiblížení matematických základů RM [13].

Uspořádaná dvojice

Uspořádaná dvojice je množina objektů s uspořádanou asociací mezi nimi. Jestliže máme objekty x a y , uspořádanou dvojici můžeme zapsat jako (x, y) nebo (y, x) . Ale výraz (x, y) je rozdílný než (y, x) [13].

Asociace

V matematice je to binární operace. To znamená, že jestliže máme více operátorů v řadě za sebou, nezáleží na pořadí, ve kterém provádíme jednotlivé operace [13].

Binární relace

Binární relace je podmnožinou kartézského součinu $X \times Y$. Uspořádaná dvojice (x, y) je binární relací z množiny X a množiny Y . Kde x je členem z množiny X a zároveň y je z množiny Y . Binární relaci pak můžeme zapsat xRy nebo $(x,y) \in R$ [13].

Kartézský součin

Množina všech uspořádaných dvojic (x, y) se nazývá kartézský součin z množiny X a Y . Kartézský součin zapisujeme $X \times Y$, kde x je z množiny X a y z množiny Y [13].

4.1.2 Relace

Nejdříve se musíme podívat na ekvivalenci. Binární relace $R \subseteq X \times X$, která je reflexivní (objekt je v relaci sám se sebou, xRx), symetrická (když xRy , potom yRx) a tranzitivní (jestliže máme relaci xRy a relaci yRz , potom platí xRz). Relace, která je reflexivní, symetrická a tranzitivní se nazývá ekvivalentní relací. Ekvivalentní třída prvku $x \in X$ závisí na všech prvcích $y \in Y$ tak jako xRy . Celou ekvivalenci R můžeme značit jako U/R [13], [14], [15], [17].

Každá podmnožina atributů CH označuje ekvivalentní relaci v U . Tato relace bude nazývána jako nerozlišitelná relace. Ekvivalence třídy je označována prvkem x a množina atributů CH se bude značit $CH(x)$. Necht' $A = (U, A)$ jako informační systém s nějakou $CH \subseteq A$ pak existuje sdružená ekvivalentní relace I_{CH} :

$$I_{CH} = \{(x,y) \in U \times U \mid \forall Attr \in CH, Attr(x) = Attr(y)\} \quad (2)$$

I_{CH} je nazývána jako CH -nerozlišitelná a budeme ji značit jako $[x]_{CH}$ [13], [14], [15], [17].

4.1.3 Aproximace množiny

Aproximace je nazývána spojitou operací, která je vnitřní a uzavřená. Předpokládáme, že existuje daná ohraničená a neprázdná množina prvků U , která se jmenuje universum a množina atributů A . Necht' X je podmnožinou U a CH je podmnožinou A . Můžeme charakterizovat množinu X pomocí atributů CH [13], [14], [15], [17].

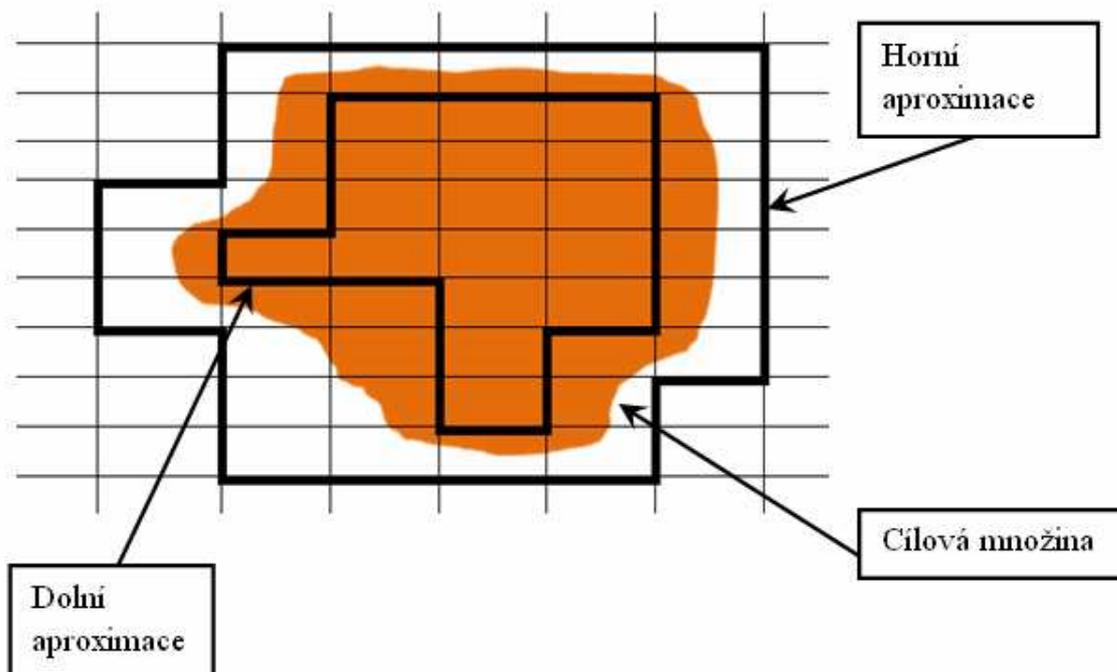
Dolní aproximace množiny X s ohledem na CH je množina všech objektů, která může být s jistotou klasifikována jako množina X , která používá atributy CH (jsou určitá X s ohledem na CH). Dolní aproximace je sjednocením všech ekvivalentních tříd $[x]_{CH}$, která je obsažená v cílové množině. Dolní aproximace je prvkem cílové množiny [13], [14], [15], [17].

Horní aproximace množiny X s ohledem na CH je množina všech objektů, která může být pravděpodobně klasifikována jako množina X , která používá atributy CH (jsou pravděpodobná X s ohledem na CH). Horní aproximace je sjednocením všech ekvivalentních tříd $[x]_{CH}$, která mají neprázdný průnik s cílovou množinou [13], [14], [15], [17].

Hraniční region množiny X s ohledem na CH je množina všech objektů, která může být vyjádřena, jako rozdíl mezi horní a dolní aproximací množiny [13], [14], [15], [17].

$$\overline{CHX} - \underline{CHX} \quad (3)$$

Dolní a horní aproximaci můžeme vidět na obrázku 4. Oranžové pole je množina atributů na mřížce ekvivalentní třídy.



Obrázek 4: Aproximace množiny [18]

Definice RM

- Množina X je **hrubá** (nepřesná s ohledem na CH), jestliže hraniční region množiny X je neprázdný.
- Množina X je **ostrá** (přesná s ohledem na CH), jestliže hraniční region množiny X je prázdný.

Dvojici $\langle \underline{CHX}, \overline{CHX} \rangle$ složené z dolní a horní aproximace říkáme RM [13], [14], [15], [17].

4.1.4 Informační systém

Při uplatňování RM jsou data obvykle prezentována v tabulce takto: Sloupce představují atributy, řádky představují objekty, a každá buňka obsahuje prvky ohodnocené příslušnými objekty a atributy. V terminologii RM jsou tyto tabulky nazývány informačním systémem.

Více formálně můžeme informační systém zapsat jako dvojici $\hat{S} = (U, A)$, U je neprázdná, ohraničená množina objektů zvaná universum a A je neprázdná, ohraničená množina atributů [13], [14], [15], [17].

4.1.5 Příklad – RM

V této podkapitole si uvedeme jednoduchý příklad RM. Celá tato podkapitola byla zpracována podle [13], [15], [17].

Informační systém

Příklad informačního systému je ukázán na tabulce. V tomto informačním systému je 10 objektů (záznamů z dopravních nehod), které jsou charakterizovány třemi atributy: Typ, věk a rodinný stav. Kde typ je reprezentován typem auta, věk se zajímá o věk pojištěného a rodinný stav představuje rodinný stav pojištěného.

Tabulka 2: Informační systém [vlastní [13]]

Objekt (Záznamy)	Atribut 1 (Typ)	Atribut 2 (Věk)	Atribut 3 (Rodinný stav)
CH1	sport	20	svobodný
CH2	sport	20	svobodný
CH3	sedan	30	svobodný
CH4	combi	30	ženatý
CH5	sedan	40	svobodný
CH6	combi	30	ženatý
CH7	sedan	30	svobodný
CH8	combi	40	rozvedený
CH9	sedan	40	svobodný
CH10	sedan	30	svobodný

Struktura ekvivalentní třídy

Jestliže plná množina atributů $A = \{Typ, Věk, Rodinný\ stav\}$ je zvážená, můžeme vidět, že máme následujících 5 ekvivalentních tříd:

$$E_1 = CH_1, CH_2,$$

$$E_2 = CH_3, CH_7, CH_{10},$$

$$E_3 = CH_4, CH_6,$$

$$E_4 = CH_5, CH_9,$$

$$E_5 = CH_8.$$

Každá ekvivalentní třída může být popsána pomocí boolean pravidla. Pro příklad E3, může být zapsán takto:

$$(Typ = combi \wedge Věk = 30 \wedge Rodinný\ stav = ženatý).$$

Rozhodovací systém

V mnohých skutečných klasifikačních problémech je výsledek již znám. Tento třídící faktor je označován jako rozhodovací atribut. Informační systém, který obsahuje rozhodovací atribut je nazýván rozhodovací systém.

Tabulka 3: Rozhodovací systém [vlastní [13]]

Objekt (Záznamy)	Atribut 1 (Typ)	Atribut 2 (Věk)	Atribut 3 (Rodinný stav)	Podvod
CH1	sport	20	svobodný	yes
CH2	sport	20	svobodný	yes
CH3	sedan	30	svobodný	yes
CH4	combi	30	ženatý	yes
CH5	sedan	40	svobodný	no
CH6	combi	30	ženatý	no
CH7	sedan	30	svobodný	no
CH8	combi	40	rozvedený	no
CH9	sedan	40	svobodný	no
CH10	sedan	30	svobodný	no

Tabulka nám ukazuje příklad rozhodovacího systému. Tato tabulka je stejná jako tabulka s výjimkou, že obsahuje rozhodovací atribut Podvod, který představuje, zda věk, pohlaví a rodinný stav mají vliv na páchání pojistného podvodu.

Aproximace množin

Příklad aproximace množin si následně ukážeme. Necht' $X = \{x | podvod(x) = Yes\}$. Z tabulky získáme $X = \{1, 2, 3, 4\}$. Dále můžeme předpokládat, že $CH = \{Typ, Věk, Rodinný\ stav\}$ pak

$$[1]_{CH} = [2]_{CH} = \{1, 2\},$$

$$[3]_{CH} = [7]_{CH} = [10]_{CH} = \{3, 7, 10\},$$

$$[4]_{CH} = [6]_{CH} = \{4, 6\},$$

$$[5]_{CH} = [9]_{CH} = \{5, 9\},$$

$$[8]_{CH} = \{8\}.$$

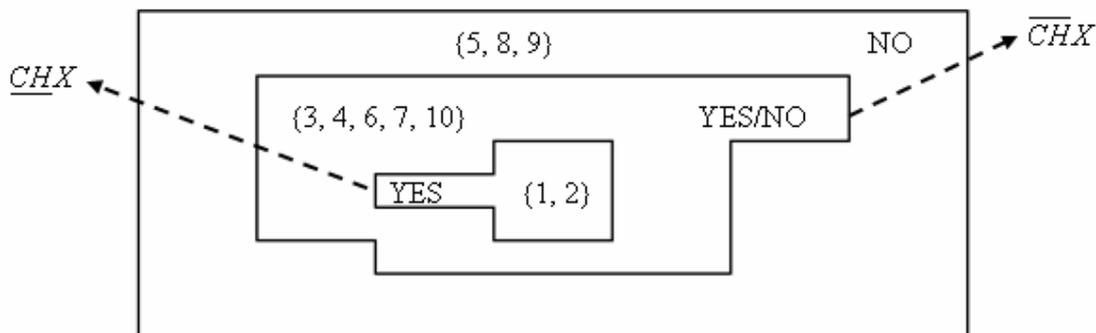
Definováním dolní a horní aproximace získáme

$$\underline{CHX} = \{1, 2\},$$

$$\overline{CHX} = \{1, 2, 3, 4, 6, 7, 10\},$$

$$\overline{CHX} - \underline{CHX} = \{3, 4, 6, 7, 10\}.$$

Vidíme, že hraniční region není prázdný a to nám ukazuje, že množina X je hrubá (rough). Objekty z množiny $\{1, 2\}$ stoprocentně patří do skupiny, které páchají pojistné podvody. Množina $\{1, 2, 3, 4, 6, 7, 10\}$ zahrnuje objekty, které pravděpodobně páchají pojistné podvody. Naopak tomu objekty 3, 4, 6, 7, 10, které jsou v hraničním regionu, nemůžou být s určitostí klasifikovány jako pojistný podvod. $U - \overline{CHX} = \{5, 8, 9\}$ nám ukazuje, že záznamy z pojistné události 5, 8, 9 určitě nevykazují známky pojistného podvodu. Aproximace množiny X můžeme vidět na obrázku 5.



Obrázek 5: Aproximace množin na příkladu pojistných podvodů [vlastní [13]]

4.2 Aplikace a softwarové nástroje

Mnoho rozdílných softwarových systémů bylo vytvořeno pomocí teorie RM. Je několik různých oblastí, ve kterých mají aplikace s využitím RM dobrý ohlas. Zde je na ukázkou několik z nich [12], [16]:

Medicína

- odhalení rakoviny prsu pomocí elektrických možností,
- klasifikace výsledků histologie,
- podpora ve výběru léčení,
- databáze lékařských rozborů,
- k odhalení pooperačních infekcí, aj.

Ekonomické, finanční a obchodní aplikace

- model pro zjednodušení manažerských kompetencí,
- předvídání finančních rizik,
- vyhodnocení rizik bankrotu,
- poskytování bankovních úvěrů,
- rozbor obchodních databází,
- vyhodnocení společností,
- chování zákazníků.

Životní prostředí

- řešení globálního oteplování,
- ochrana životního prostředí,
- stabilizace globální teploty,
- programování systému dodávky vody,
- kontrolování podmínek v suchých nádržích.

Inženýrství

- zvýšení audio signálů,
- rozeznání hudebních zvuků,
- rozbor obrazů,
- filtrace a kódování zvuků,

- inteligentní plánování,
- měření akustiky koncertních hal,
- redukce hluků a zkreslení v digitálních zvukových signálech,
- technická diagnostika mechanických objektů,
- rozbor materiálů.

Informační vědy

- ohodnocování kvality softwarů,
- získávání informací,
- RM model relačních databází,
- data mining pro hudební databáze.

Sociální oblast a ostatní

- rozbor konfliktů,
- studie volebních preferencí,
- on-line předpověď postupu hry při zápasu ve volejbalu,
- výběr sociálních funkcí.

Molekulární biologie

- objevování použitelných složek bílkovin ze sledu aminokyselin.

Farmacie

- rozbor souvislostí mezi strukturou a aktivitou substancí.

4.2.1 Softwarové nástroje

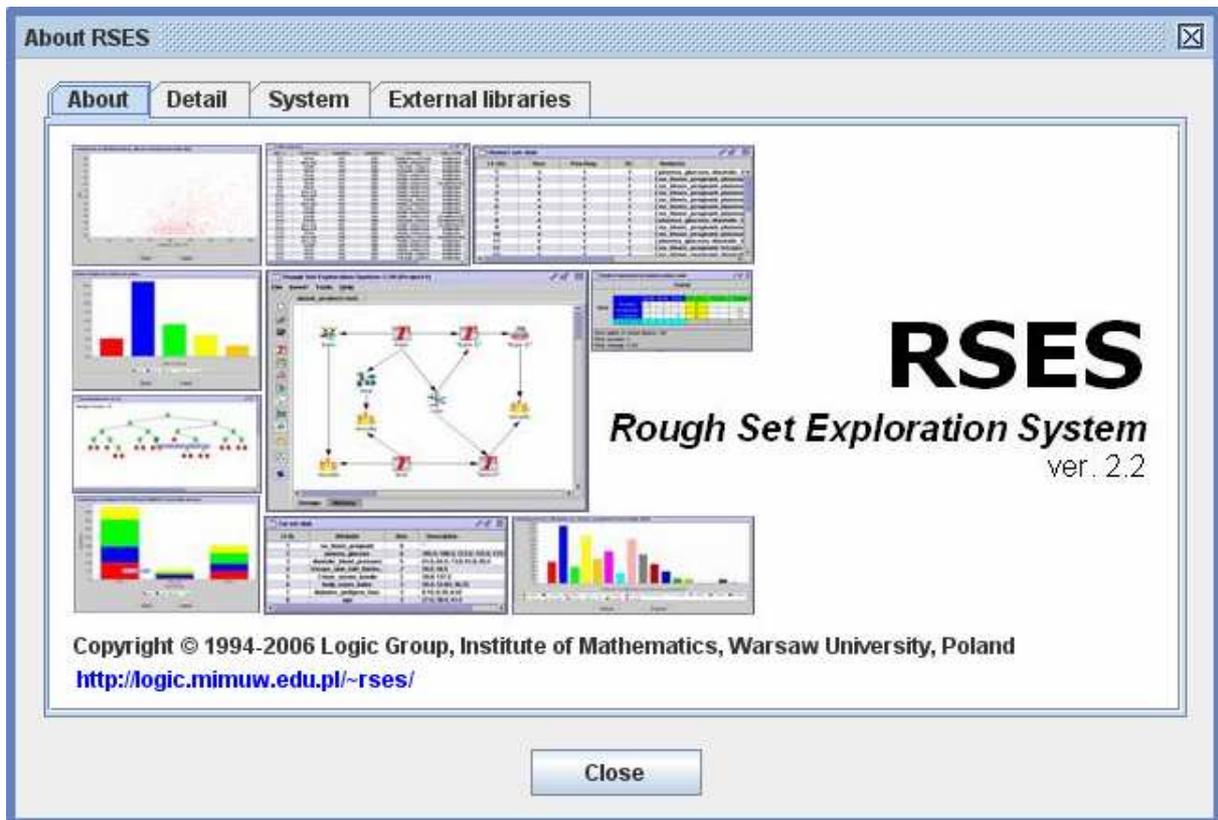
Existuje mnoho systémů používajících RM pro rozbor dat. Různé produkty jsou volně ke stažení, jiné jsou pouze pro komerční účely. Některé z nich jsou univerzální, jako např. RSES2, ROSETTA, ARES. A další jsou specializovaná, více zaměřené na cíl, jako jsou LERS, PRIMEROSE (pouze obsahující pravidla), RoughFuzzyLab (úmyslně vytvořené na rozpoznávání obrazů) [18].

Představme si softwarový produkt, který pracuje na podkladě RM.

Rough Set Exploration System (RSES2)

RSES je nástroj pro analýzu dat v tabulkách, tak jako ukazuje obrázek 6. Je založen na základních metodách a algoritmech, která jsou z oblasti RM. Skládá se ze dvou hlavních součástí – grafické uživatelské rozhraní front-end a výpočtového jádra.

RSES byl vytvořen a implementován jako výsledek průzkumu hrubých množin (RM). Tento software byl vyvinut Andrejem Skowronem a jeho týmem na Varšavské Univerzitě v Polsku [19].



Obrázek 6: RSES2 [19]

Další softwarové nástroje

Zde jsou uvedeny další softwarové produkty, které pracují na základě RM:

- Datalogic/R, <http://ourworld.compuserve.com/homepages/reduct>,
- Grobian (Roughian), e-mail: I.Duentsch@ulst.ac.uk, ggediga@luce.psychology.Uni-Osnabrueck.DE,
- LERS-A Knowledge Discovery System , e-mail: jerzy@eecs.ukans.edu,
- PRIMEROSE, e-mail: tsumoto@computer.org,
- ProbRough | A System for Probabilistic Rough Classifiers Generation, e-mail: {zpiasta,lenarcik}@sabat.tu.kielce.pl,

- Rough Family - Software Implementation of the Rough Set Theory, e-mail: Roman.Slowinski@cs.put.poznan.pl, Jerzy.Stefanowski@cs.put.poznan.pl,
- RSDM: Rough Sets Data Miner, e-mail: {cfbaizan, emenasalvas}@.fi.upm.es,
- RoughFuzzyLab - a System for Data Mining and Rough and Fuzzy Sets Based Classification, e-mail: rswiniar@saturn.sdsu.edu,
- RSL - The Rough Set Library, <ftp://ftp.ii.pw.edu.pl/pub/Rough/>,
- TAS: Tools for Analysis and Synthesis of Concurrent Processes using Rough Set Methods, e-mail: zsuraj@univ.rzeszow.pl,
- Trance: a Tool for Rough Data Analysis, Classification, and Clustering, e-mail:wojtek@cs.vu.nl

5 POUŽITÍ RM VE FORENZNÍM AUDITU

V této části si na praktické ukázce dat přiblížíme použití RM pro odhalení pojistných podvodů. Data pro ukázkou byla získána jako skutečné pojistné události, které nastaly v jedné australské pojišťovně v roce 1994 až 1996 v oblasti pojištění zodpovědnosti za škodu způsobenou jízdou motorového vozidla a havarijního pojištění. K dispozici je 15 419 záznamů, z kterých jen 6 % tvoří záznamy se známkou podvodu.

Při navrhování modelu se postupovalo podle metodiky CRISP-DM, která byla navržena jako univerzální postup pro dobývání znalostí z databází. Tento standard je popsán jako hierarchický model, který je podle této metodiky složen z šesti fází [20]:

Porozumění problematice

V této úvodní fázi se zaměřujeme na pochopení potřeb projektu, jeho cílů, požadavků a následovní převedení těchto znalostí do zadání úlohy pro dobývání znalostí z databází. Dále se v této fázi stanovuje předběžný plán prací.

Porozumění datům

Tato fáze začíná sběrem potřebných dat, pokračuje prvotním seznámením se s daty, souvislostmi, jejich kvalitou, vyhledáváním zajímavých podmnožin. Vyslovujeme první hypotézy o neznámých znalostech.

Příprava dat

Příprava dat je obvykle nejpracnější částí celého projektu. Zaobírá se čištěním dat, zlepšováním jejich kvality a transformací dat. Touto fází se bude podrobně věnovat samostatná kapitola této práce.

Modelování

V této fázi se na data aplikují rozdílné metody a techniky. Touto fází se taktéž budeme podrobně věnovat.

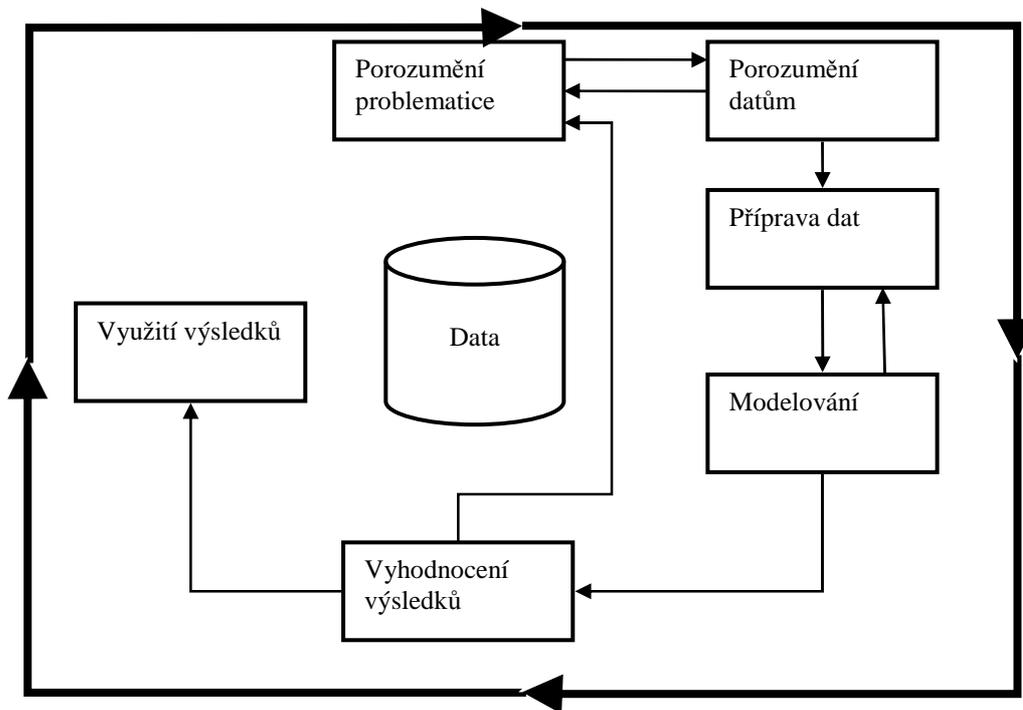
Vyhodnocení výsledků

Vyhodnocení úspěšnosti metody se budeme věnovat také v samostatné kapitole. V této fázi se dosažené výsledky vyhodnocují z pohledu manažerů. Rozhoduje se o způsobu využití výsledků.

Využití výsledků

Vytvoření modelu neznamena konec projektu. I v případě, že cílem bylo lepší pochopení dat, je potřeba získané znalosti upravit, uspořádat a zorganizovat.

Pořadí jednotlivých kroků není pevně dáno. Jak je dáno na obrázku 7 v rámci této metodiky je možná zpětná vazba, přehodnocení jednotlivých kroků, jejich přepracování atd.



Obrázek 7: Metodika CRISP-DM [20]

5.1 Příprava dat

Z časového hlediska je tento krok projektu nejpracnější a nejnáročnější. Uvádí se, že fáze předzpracování dat zabírá až 80 % času. Zároveň je to fáze, která má rozhodující význam pro úspěch dané aplikace [18].

Databáze v dnešních organizacích a jejich využití pro vytažení znalostí přinášejí mnoho problémů. První z nich bývá nekompletnost údajů (chybějí údaje pro konkrétní atributy, případně atributy vhodné na zkoumání se v databázi nenacházejí). Existuje několik důvodů existence těchto nedostatků. Může to být způsobeno nekvalitními nástroji pro sběr dat, nebo uživatelskými chybami při jejich zadávání. Nástroje pro čištění dat jsou schopné doplňovat chybějící data. Mezi nejpoužívanější metody čištění dat patří v případě chybějících hodnot následující [20]:

- ignorace záznamu s chybějící hodnotou,
- nahrazení chybějící hodnoty hodnotou „nevím“ (může být časově náročné),
- místo chybějící hodnoty použít konstantu,
- manuální doplnění chybějící hodnoty (často téměř nemožné),
- nahrazení chybějícího záznamu nejčtenější hodnotou z existujících hodnot atributu,
- atribut je nahrazený střední hodnotou, průměrem, nejpravděpodobnější hodnotou, proporcionálním podílem všech hodnot, atd.

Dalším problémem, se kterým je potřeba se vypořádat, je velikost databáze. Ve většině případů se jedná o databáze, které obsahují desítky tisíc objektů (někdy se může jednat o miliony či miliardy záznamů). V případě, že algoritmy pracují v dávkovém režimu, kde se předpokládá zpracovat všechna trénovací data naráz v operační paměti, nám může vzniknout s příliš velkou databází problém. V tomto případě se doporučuje tzv. selekce. To znamená, že použijeme jen určitý vzorek vybraný ze všech dat [20].

5.1.1 Vlastní příprava dat

Data v této diplomové práci byla získána jako volný textový soubor. Tento soubor obsahuje 15 419 záznamů a 33 atributů. Tento textový soubor bylo potřebné přetransformovat do formátu vhodného k použití v RSES2. Data byla zpracována pomocí tabulkového editoru MS Excel.

Seznam atributů

K dispozici je velké množství atributů popisující každou pojistnou událost. V následující tabulce 4 je seznam všech použitých atributů. Tabulka obsahuje název proměnné, její popis a datový typ.

Tabulka 4: Seznam atributů[zdroj: vlastní]

Název atributu	Popis	Datový typ
Měsíc	měsíc, v kterém nastala pojistná událost	řetězec
Týden v měsíci	pořadí týdne v měsíci, v kterém nastala pojistná událost	číslo
Den týdne	den, v kterém nastala pojistná událost	řetězec
Značka	značka automobilu na pojistné smlouvě	řetězec

Místo nehody	místo nehody (město, mimo město)	boolean
Den ohlášení	den, v kterém byla ohlášena pojistná událost	řetězec
Měsíc ohlášení	měsíc, v kterém byla ohlášena pojistná událost	řetězec
Týden ohlášení	pořadí týdne v měsíci, v kterém byla ohlášena pojistná událost	číslo
Pohlaví	pohlaví pojištěného	boolean
Rodinný stav	rodinný stav pojištěného	řetězec
Věk	věk pojištěného	číslo
Zodpovědnost	na čí straně byla zodpovědnost za nehodu (pojištěný, třetí strana)	boolean
Typ pojištění	typ pojištění (odpovědnost za škodu, havarijní)	řetězec
Typ auta	typ vozidla (sport, sedan, nákladní auto, ...)	řetězec
Cena vozidla	cena vozidla	číslo
Číslo pojištění	číslo pojistné události	číslo
Číslo reprezentanta	číslo reprezentanta v pojišťovně vyřizujícího pojistnou událost	číslo
Pojistná suma	pojistná suma	číslo
Rating řidiče	rating řidiče, hodnocení od 1 do 5	číslo
Délka pojištění	délka pojištění v dnech	číslo
Délka nahlášení	dny uplynulé od nastání pojistné události na dané pojistné smlouvě	číslo
Počet předchozích událostí	počet předcházejících pojistných událostí	číslo
Věk vozidla	věk vozidla v letech	číslo
Věk pojištěného - interval	věk pojištěného - interval	řetězec
Hlášení	zda bylo vyplněné policejní hlášení (ano, ne)	boolean
Svědci	zda měla pojistná událost svědky	boolean

Typ agenta	typ agenta (externí, interní)	boolean
Počet připojištění	počet připojištění na dané pojistné smlouvě	číslo
Změna adresy	zda pojištěný před nehodou měnil adresu trvalého pobytu	řetězec
Počet vozů	počet vozidel ve vlastnictví pojištěného	číslo
Rok	rok, v kterém nastala pojistná událost	číslo
Základní pojistka	typ základní pojistky	řetězec
Podvod	zda se jedná o podvod (ano, ne)	boolean

Pro další postup bylo nutné tyto data upravit, a to tak, že se text nahradil číslem. Změněné atributy, které použijeme pro modelaci v RSES2, vypadají takto:

„Attr0“

Tento atribut znamená měsíc, ve kterém nastala pojistná událost. Modifikováno je dle následující tabulky 5.

Tabulka 5: Attr0 [zdroj: vlastní]

Jan	1
Feb	2
Mar	3
Apr	4
May	5
Jun	6
Jul	7
Aug	8
Sep	9
Oct	10
Nov	11
Dec	12

„Attr1“

Pořadí týdne v měsíci, v kterém nastala pojistná událost. Tato vlastnost je v tabulce nezměněná.

„Attr2“

Den, ve kterém nastala pojistná událost. Upraveno dle tabulky 6

Tabulka 6: Attr2 [zdroj: vlastní]

Monday	1
Tuesday	2
Wednesday	3
Thursday	4
Friday	5
Saturday	6
Sunday	7

„Attr3“

Značka automobilu na pojistné smlouvě, poupraveno podle tabulky 7.

Tabulka 7: Attr3 [zdroj: vlastní]

Accura	1
BMW	2
Dodge	3
Ferrari	4
Ford	5
Honda	6
Chevrolet	7
Jaguar	8
Lexus	9
Mazda	10
Mercedes	11
Mercury	12
Nissan	13
Pontiac	14
Porsche	15
Saab	16
Saturn	17
Toyota	18
VW	19

„Attr4“

Místo nehody je datový typ boolean. Dopravní nehoda se mohla stát ve městě – 1, nebo mimo město – 0.

„Attr5“

Den, ve kterém byla nahlášena pojistná událost. Změněno stejně jako „attr2“.

„Attr6“

Tato charakteristika tabulky znamená měsíc, ve kterém byla nahlášená pojistná událost.

Atribut je zpracován totožně s „attr0“.

„Attr7“

Pořadí týdne v měsíci, ve kterém byla ohlášena pojistná událost, je beze změny.

„Attr8“

Pohlaví pojištěného, muž – 1, ženě je přiřazena hodnota 0.

„Attr9“

Rodinný stav pojištěného, přepracováno na tabulku 8.

Tabulka 8: Attr9 [zdroj: vlastní]

Divorced	1
Single	2
Married	3
Widow	4

„Attr10“

Věk pojištěného, tato vlastnost zpracována nezměněná.

„Attr11“

Na které straně byla v době nehody zodpovědnost (pojištěný – 1, třetí strana – 0).

„Attr12“

Typ pojištění (sedan, sport, nákladní auto, havarijní pojištění, odpovědnost za škodu, havarijní + odpovědnost) viz tabulka 9.

Tabulka 9: Attr12 [zdroj: vlastní]

Sedan - All Perils	1
Sedan - Collision	2
Sedan - Liability	3
Sport - All Perils	4
Sport - Collision	5
Sport - Liability	6
Utility - All Perils	7
Utility - Collision	8
Utility - Liability	9

„Attr13“

Typ vozidla tak jako v tabulce 10.

Tabulka 10: Attr13 [zdroj: vlastní]

Sedan	1
Sport	2

Utility	3
---------	---

„Attr14“

Cena vozidla změněno podle intervalu v tabulce 11.

Tabulka 11: Attr14 [zdroj: vlastní]

less than 20,000	1
20,000 to 29,000	2
30,000 to 39,000	3
40,000 to 59,000	4
60,000 to 69,000	5
more than 69,000	6

„Attr15“

Tento atribut – číslo pojistné události je ponechán beze změny.

„Attr16“

Číslo reprezentanta v pojišťovně vyřizujícího pojistnou událost taktéž nezměněno.

„Attr17“

Pojistná suma, upraveno dle tabulky 12

Tabulka 12: Attr17 [zdroj: vlastní]

300	1
400	2
500	3
600	4
700	5

„Attr18“

Rating řidiče se pohybuje v rozmezí 1 (nejlepší řidič) až po rating 5 (nejhorší řidič), taktéž beze změny.

„Attr19“

Délka pojištění ve dnech viz tabulka 13.

Tabulka 13: Attr19 [zdroj: vlastní]

none	1
1 to 7	2
8 to 15	3
15 to 30	4
more than 30	5

„Attr20“

Dny uplynuté od nastání pojistné události po její nahlášení, upraveno stejně jako „attr19“.

„Attr21“

Počet předcházejících pojistných událostí na dané pojistné smlouvě, změněno podle této tabulky 14.

Tabulka 14: Attr21 [zdroj: vlastní]

None + 1	1
2 - 4	2
more 4	3

„Attr22“

Věk vozidla v letech, použito s pomocí takto upravené tabulky 15.

Tabulka 15: Attr22 [zdroj: vlastní]

new	1
2 years	2
3 years	3
4 years	4
5 years	5
6 years	6
7 years	7
more than 7	8

„Attr23“

Věk pojištěného – rozděleného na intervaly do osmi skupin dle tabulky 16.

Tabulka 16: Attr23 [zdroj: vlastní]

16 to 17	1
18 to 20	2
21 to 25	3
26 to 30	4
31 to 35	5
36 to 40	6
41 to 50	7
51 to 65	8
over 65	9

„Attr24“

Příznak, zda bylo vyplněno policejní hlášení či nikoliv (ano – 0, ne – 1).

„Attr25“

Pojistná událost se stala za přítomnosti svědků, či bez nich. Typu boolean (byli svědci – 0, nebyli svědci – 1).

„Attr26“

Typ agenta (externí – 1, interní – 0).

„Attr27“

Počet připojištění na dané pojistné smlouvě, dle tabulky 17.

Tabulka 17: Attr27 [zdroj: vlastní]

none	1
1 to 2	2
3 to 5	3
more than 5	4

„Attr28“

Znak toho, zda pojištěný před nehodou změnil adresu trvalého pobytu, či nikoliv. Zpracováno dle tabulky 18.

Tabulka 18: Attr28 [zdroj: vlastní]

under 6 months	1
1 year	2
2 to 3 years	3
4 to 8 years	4
no change	5

„Attr29“

Celkový počet aut, která vlastní pojištěný. Použito pomocí tabulky 19.

Tabulka 19: Attr29 [zdroj: vlastní]

1 vehicle	1
2 vehicles	2
3 to 4	3
5 to 8	4
more than 8	5

„Attr30“

Rok, ve kterém nastala pojistná událost (rok 1994 – 1, rok 1995 – 2, rok 1996 -3).

„Attr31“

Typ základní pojistky, zda se jedná o havarijní, odpovědnost, nebo kombinace obou. Zpracováno podle tabulky 20.

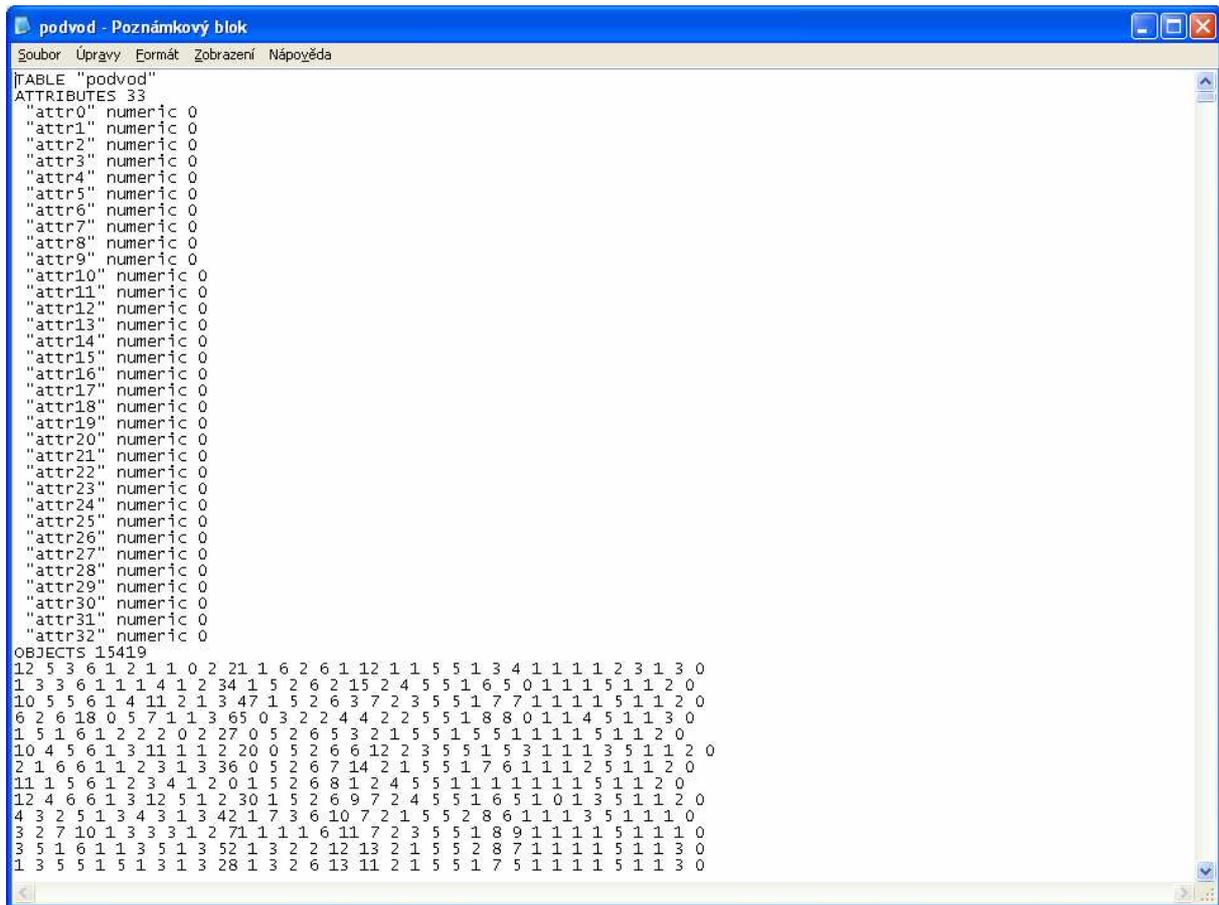
Tabulka 20: Attr31 [zdroj: vlastní]

All Perils	1
Collision	2
Liability	3

„Attr32“

Faktor, který poukazuje na pojistný podvod, anebo ne (ano – 1, ne – 0).

Připravené atributy pro modelaci v RSES2 se do softwaru vkládají z poznámkového bloku a mají tento formát, viz obrázek 8



Obrázek 8: upravená data [zdroj: vlastní]

Protože použitá data obsahují příliš mnoho objektů, byla použita selekce těchto dat. Aplikovaný vzorek dat má pro naši potřebu 300 objektů a vybrány byly pomocí náhodného výběru. Takto vybrané a upravené atributy se použily v dalším kroku – modelování.

5.2 Modelování

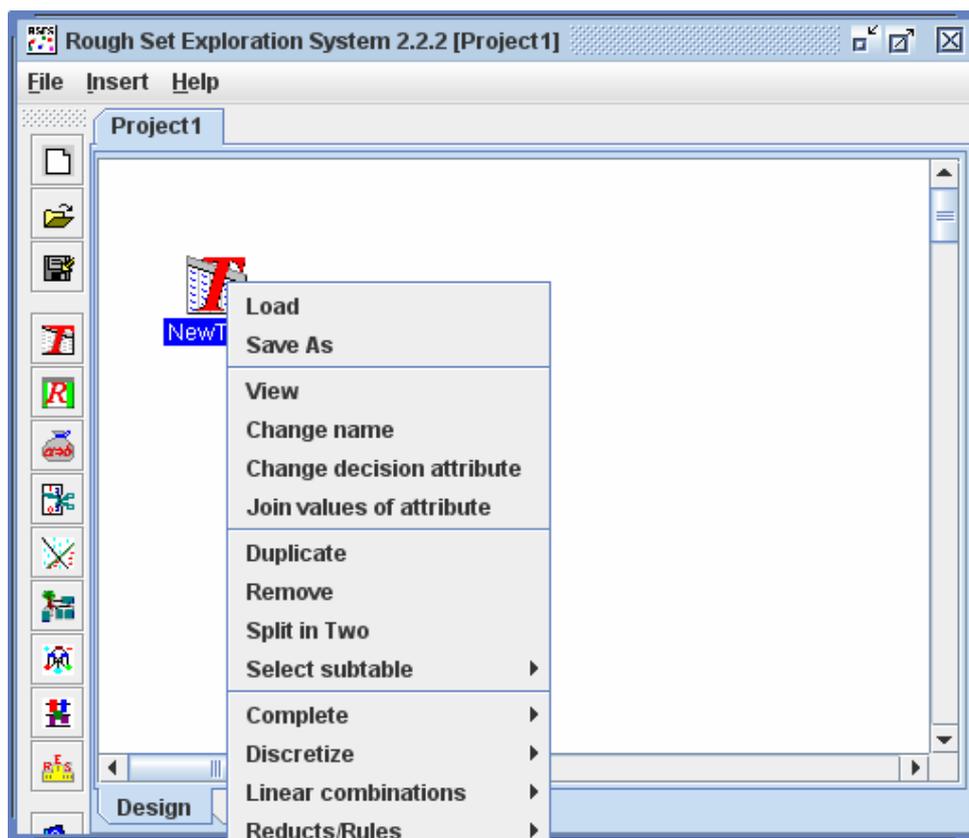
V této části diplomové práce vytvoříme model pro odhalení pojistných podvodů. Tento krok byl proveden v softwaru Rough Set Exploration System 2.2.2 (RSES2), který je založen na základě RM.

5.2.1 Vlastní postup modelování

Máme k dispozici dostatek atributů, proto existuje předpoklad, že by model mohl být přínosný a pojišťovně by mohl ušetřit část nákladů na výplatu plnění.

Vložení upravených dat

Pomocí nabídky insert table, vložíme novou tabulku. Data do tabulky načteme pomocí pravého tlačítka myši a zvolíme Load, viz obrázek 9. V našem příkladu vybereme soubor podvod.tab.



Obrázek 9: načtení dat [zdroj: vlastní]

Načtenou tabulku, v našem případě je to informační a rozhodovací systém, můžeme prohlédnout pomocí View. Na následujícím obrázku 10 můžeme vidět, jakou má naše tabulka podobu.

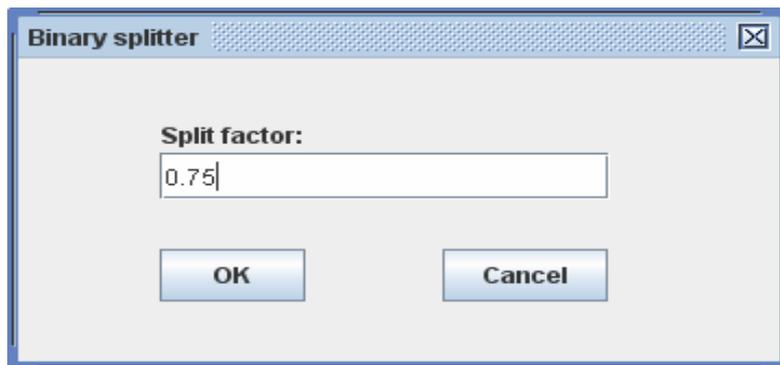
300 / 33	attr0	attr1	attr2	attr3	attr4
O:1	12	5	3	6	1
O:2	1	3	3	6	1
O:3	10	5	5	6	1
O:4	6	2	6	18	0
O:5	1	5	1	6	1
O:6	10	4	5	6	1
O:7	2	1	6	6	1
O:8	11	1	5	6	1
O:9	12	4	6	6	1
O:10	4	3	2	5	1
O:11	3	2	7	10	1
O:12	3	5	1	6	1
O:13	1	3	5	5	1
O:14	1	5	5	6	0
O:15	1	5	1	5	1
O:16	8	4	2	5	1
O:17	4	4	4	5	1
O:18	7	5	7	7	1
O:19	5	4	4	14	1
O:20	4	4	1	6	1
O:21	4	2	5	10	1
O:22	1	2	6	7	1
O:23	8	3	7	10	1
O:24	6	3	6	14	1
O:25	9	3	5	10	1

Obrázek 10: Informační a rozhodovací systém[zdroj: vlastní]

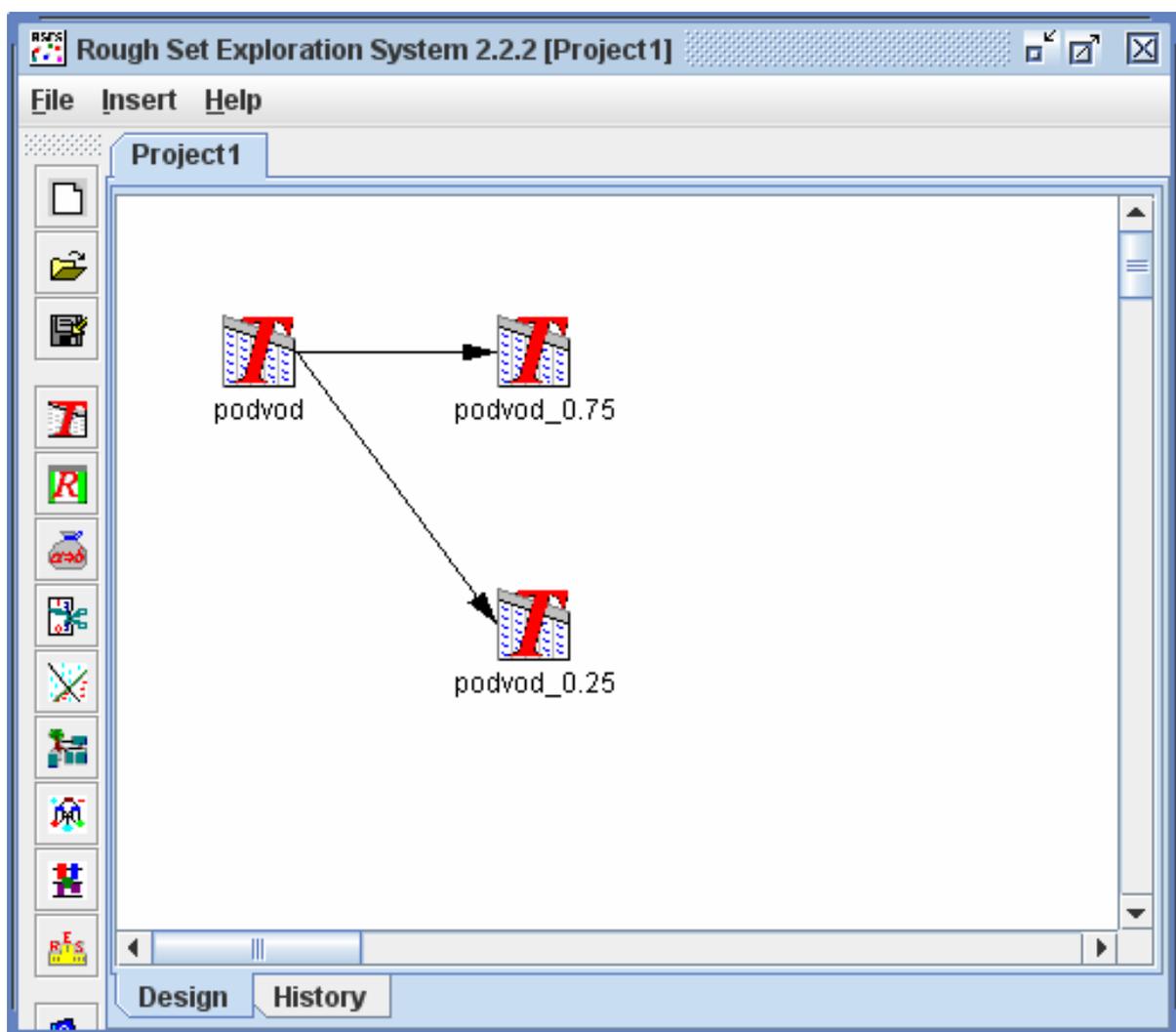
Rozdělení dat na trénovací a testovací množinu

V případě, že již máme tabulku dat, ze kterých chceme provést klasifikaci a predikci, musíme nejdříve data rozdělit na dvě disjunktní části. Z jedné části dat se stane trénovací vzorek, druhá část bude použita jako testovací množina pro klasifikaci sestavenou z trénovacích dat [20].

Data rozdělíme, podle metody Hold-out, v poměru 75 % (trénovací data) a 25 % (testovací data). Jde o empiricky doporučený poměr trénovacích a testovacích dat. Pro vytvoření těchto množin vybereme Split in two a doplníme hodnotu 0,75 jako Split factor viz obrázek 11. Výsledek můžeme spatřit na obrázku 12 [20].



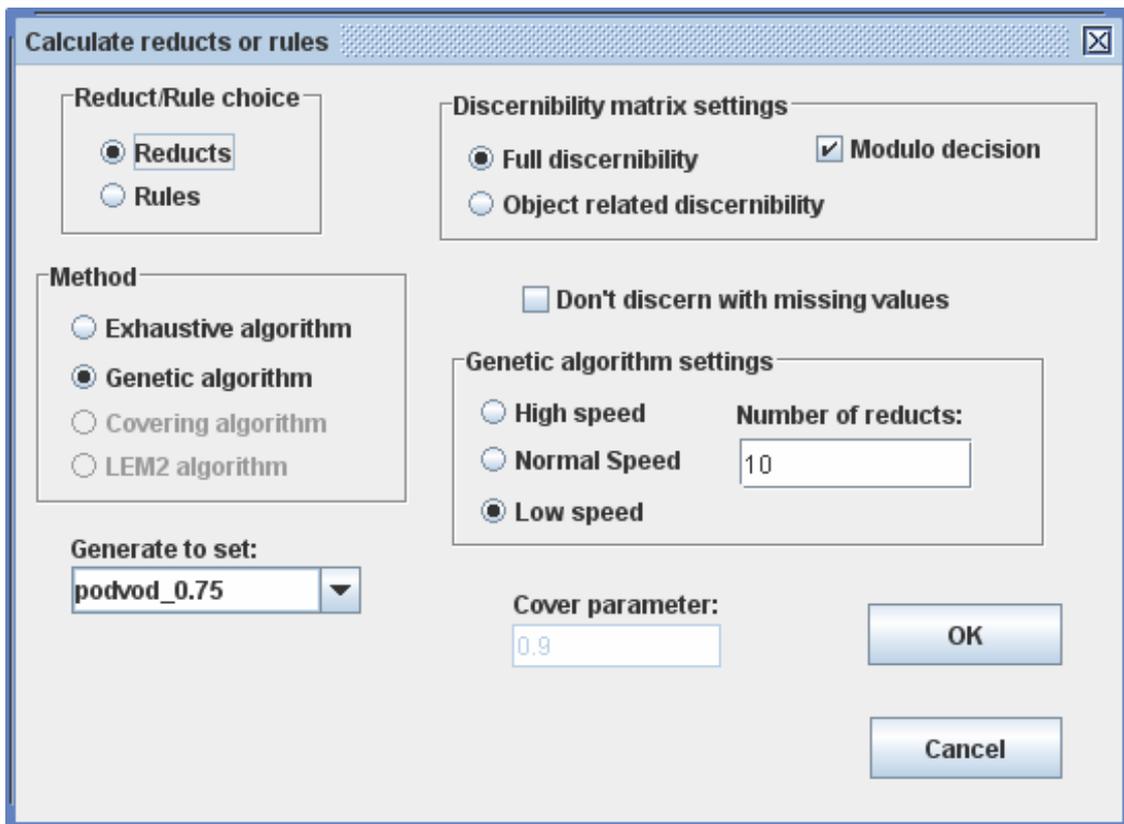
Obrázek 11: split factor [zdroj: vlastní]



Obrázek 12: rozdělení dat na trénovací a testovací množinu [zdroj: vlastní]

Redukce dat a generování pravidel

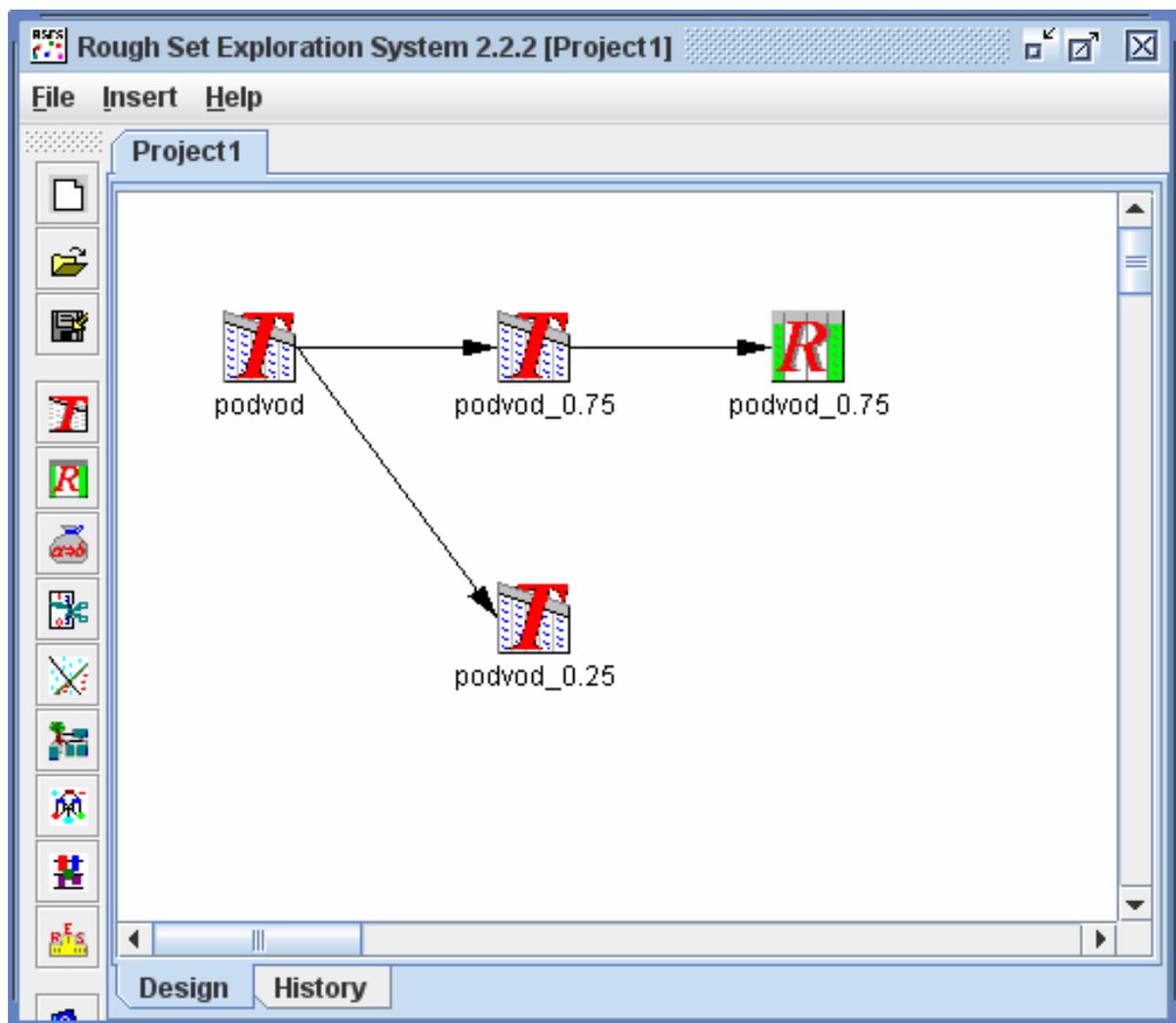
V dalším kroku trénovací data (podvod_0.75) zredukujeme a následně z výsledné redukce vygenerujeme rozhodovací pravidla. Pro tento účel vybereme z nabídky Reducts/Rules/Calculate reducts or rules. Pro redukci dat zvolíme metodu Genetic algorithm a jako rychlost označíme Low speed, počet ponecháme na deseti dle obrázku 13. Zredukovanou množinu můžeme vidět na obrázku 14 a obrázku 15.



Obrázek 13: Redukce dat 1 [zdroj: vlastní]

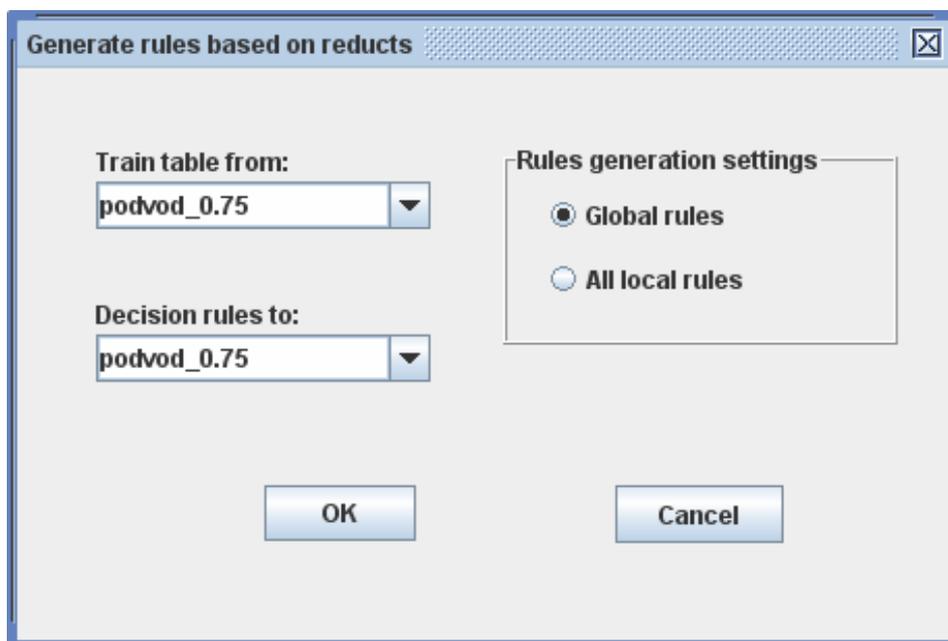
(1-10)	Size	Pos.Reg.	SC	Reducts
1	1	1	1	{ attr15 }
2	3	1	1	{ attr1, attr2, attr10 }
3	3	1	1	{ attr0, attr2, attr10 }
4	3	1	1	{ attr0, attr1, attr10 }
5	3	1	1	{ attr10, attr16, attr18 }
6	3	1	1	{ attr5, attr10, attr16 }
7	3	1	1	{ attr2, attr5, attr10 }
8	4	1	1	{ attr1, attr3, attr16, attr30 }
9	4	1	1	{ attr2, attr6, attr10, attr14 }
10	4	1	1	{ attr2, attr7, attr16, attr30 }

Obrázek 14: Redukovaná množina [zdroj: vlastní]



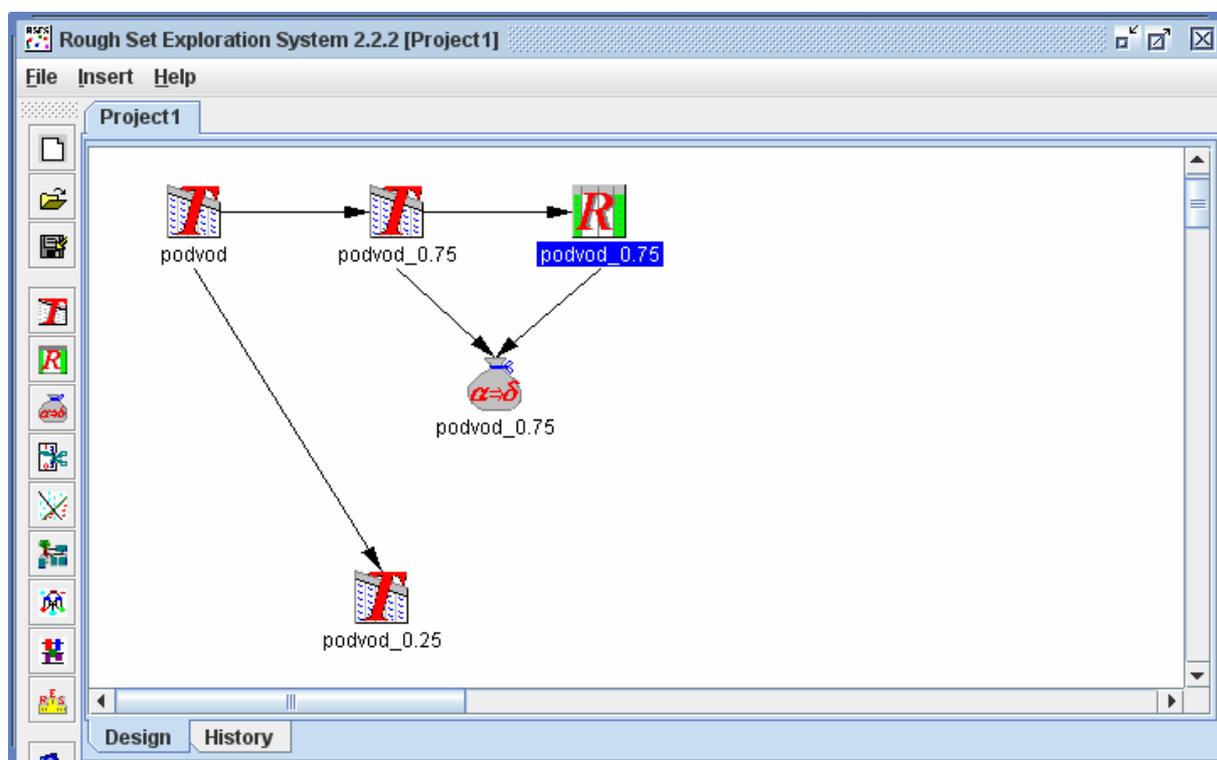
Obrázek 15: Redukce dat 2 [zdroj: vlastní]

Dalším krokem je vygenerování pravidel. Tato pravidla vygenerujeme z již zredukované množiny pomocí nabídky Generate rules, jak můžeme vidět na obrázku 16.



Obrázek 16: Generování pravidel [zdroj: vlastní]

Výsledná pravidla můžeme vidět na obrázku 17 a obrázku 18.



Obrázek 17: Generovaná pravidla 1 [zdroj: vlastní]

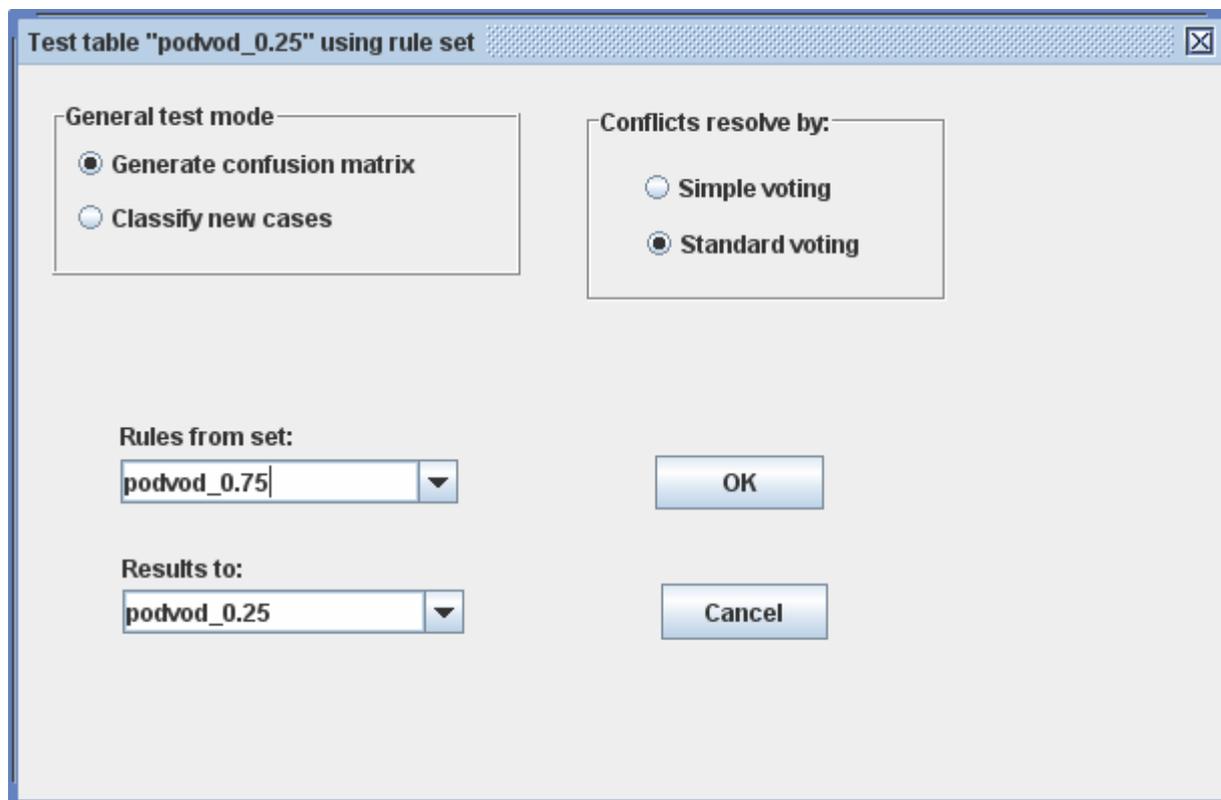
(1-214...	Match	Decision rules
1	1	(attr 15=1)=>(attr 32={0[1D])
2	1	(attr 15=2)=>(attr 32={0[1D])
3	1	(attr 15=3)=>(attr 32={0[1D])
4	1	(attr 15=5)=>(attr 32={0[1D])
5	1	(attr 15=6)=>(attr 32={0[1D])
6	1	(attr 15=7)=>(attr 32={0[1D])
7	1	(attr 15=8)=>(attr 32={0[1D])
8	1	(attr 15=9)=>(attr 32={0[1D])
9	1	(attr 15=11)=>(attr 32={0[1D])
10	1	(attr 15=12)=>(attr 32={0[1D])
11	1	(attr 15=13)=>(attr 32={0[1D])
12	1	(attr 15=15)=>(attr 32={0[1D])
13	1	(attr 15=16)=>(attr 32={0[1D])
14	1	(attr 15=17)=>(attr 32={0[1D])
15	1	(attr 15=18)=>(attr 32={0[1D])
16	1	(attr 15=19)=>(attr 32={0[1D])
17	1	(attr 15=20)=>(attr 32={0[1D])
18	1	(attr 15=25)=>(attr 32={0[1D])
19	1	(attr 15=26)=>(attr 32={0[1D])
20	1	(attr 15=28)=>(attr 32={0[1D])
21	1	(attr 15=29)=>(attr 32={1[1D])
22	1	(attr 15=30)=>(attr 32={0[1D])
23	1	(attr 15=31)=>(attr 32={0[1D])
24	1	(attr 15=32)=>(attr 32={0[1D])
25	1	(attr 15=33)=>(attr 32={0[1D])
26	1	(attr 15=34)=>(attr 32={0[1D])
27	1	(attr 15=37)=>(attr 32={0[1D])
28	1	(attr 15=38)=>(attr 32={0[1D])
29	1	(attr 15=39)=>(attr 32={0[1D])
30	1	(attr 15=41)=>(attr 32={0[1D])
31	1	(attr 15=42)=>(attr 32={0[1D])

Obrázek 18: Generovaná pravidla 2 [zdroj: vlastní]

Testování na testovacích datech

Posledním krokem je provést testování na testovacích datech. Cílem našeho testování je určit v kolika případech se klasifikátor shoduje s učitelem a v kolika případech se dopustil chyb. Testovat pravidla budeme pomocí testovací množiny (podvod_0.25).

Tyto výsledky je obvyklé zachytit v matici záměn (confusion matrix). Pro vložení výsledku vybereme z menu Classify/Test table using rule set viz obrázek 19. Přes příkaz View si otevřeme matici záměn.



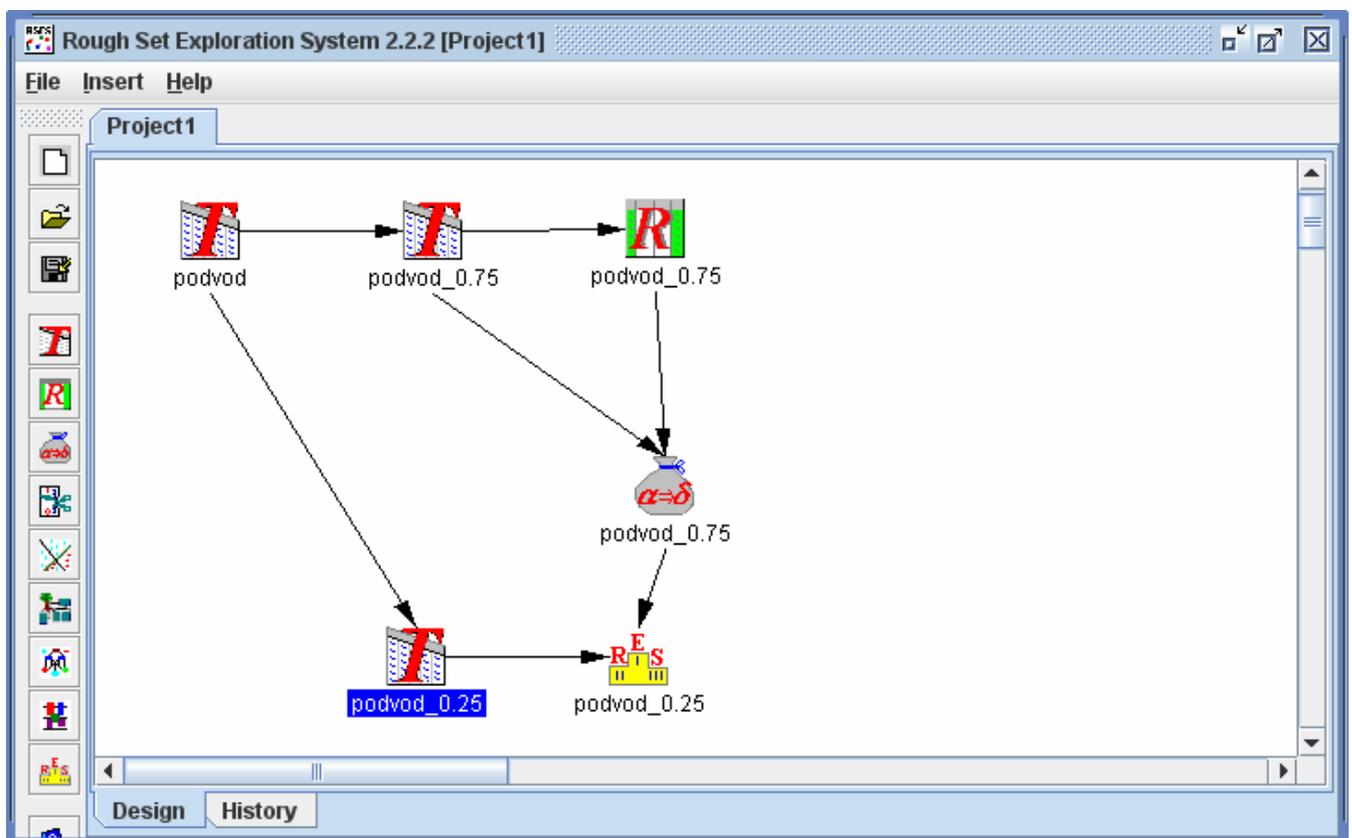
Obrázek 19: Testování na testovacích datech [zdroj: vlastní]

Ve sloupcích jsou uvedeny informace, jak RSES2 postupoval při klasifikaci. V řádcích pak nalezneme informace, jak by to mělo být podle učitele. Matice záměn pozoruje pouze počty správně a nesprávně zařazených příkladů. [20]

Celková správnost, která se nazývá i úspěšnost, je nejjednodušší charakteristika toho, jak jsou získané znalosti kvalitní. Vidíme, že celková správnost (totaly accuracy) je 96,9 %, tzn., že data jsou zatříděna správně k počtu testovaných případů, dle obrázek 20. Celkový model můžeme vidět na obrázku 21. [20]

Results of experiments by train&test method: podvod_0.25						
		Predicted				
Actual		0	1	No. of obj.	Accuracy	Coverage
	0	31	1	74	0.969	0.432
	1	0	0	1	0	0
True positive rate		1	0			
Total number of tested objects: 75						
Total accuracy: 0.969						
Total coverage: 0.427						

Obrázek 20: Matice záměn [zdroj: vlastní]



Obrázek 21: Výsledný model [zdroj: vlastní]

Závěr modelování

Tento experiment byl proveden opakovaně a to při použití různých metod a nastavení těchto metod. Provedené experimenty jsou součástí přílohy 1.

Nejlepších výsledků se dosahovalo při rozdělení trénovacích a testovacích dat, metodou Hold-out v poměru 75 % ku 25 %. Při jiném rozdělení vycházely horší výsledky, s celkovou správností kolem 90 %, které by za normálních okolností byly brány jako ucházející.

Při redukci dat dosahoval experiment nejhorších výsledků při výběru vysoké rychlosti a vysokého počtu zredukovaných množin. Celková správnost se pohybovala kolem 85 %, kterou bychom taktéž mohli považovat za ucházející. Při zvolení nízké rychlosti a nižšího počtu zredukovaných množin, vycházel experiment překvapivě s velmi dobrými výsledky, které jsou znázorněny v předchozím příkladu.

Vzhledem k nízké kapacitě použitelného softwarového produktu, musela být u dat použitých v experimentu provedena selekce pomocí náhodného výběru na 300 objektů. I přesto byl použitý vzorek dat přijatelný pro daný experiment.

5.3 Výstupy z modelování

Závěrem by se dalo říci, že použitelnost modelu, viz obrázek 22, by mohl pomoci k detekci pojistných podvodů. Při vyhodnocování úspěšnosti musíme brát v úvahu i náklady spojené s náhradou a pojistným plněním, tak jako s vyšetřováním běžných i podezřelých pojistných událostí.

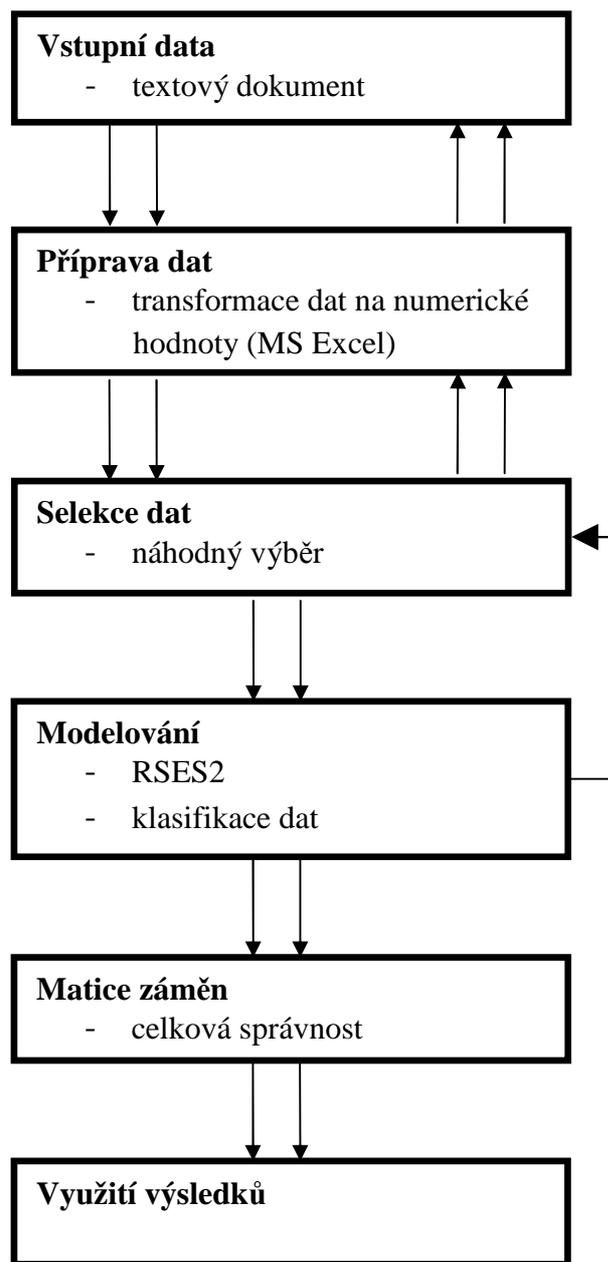
Model bude úspěšný i v případě, když finanční prostředky nebudou ušetřeny, protože musíme vzít do úvahy image pojišťovny a její reputaci mezi běžnými zájemci.

Případné odhalení pojistných podvodů, totiž odradí možné zájemce o jeho spáchání v dané pojišťovně. Toto může do budoucnosti pojišťovně taktéž ušetřit nemalou část finančních prostředků.

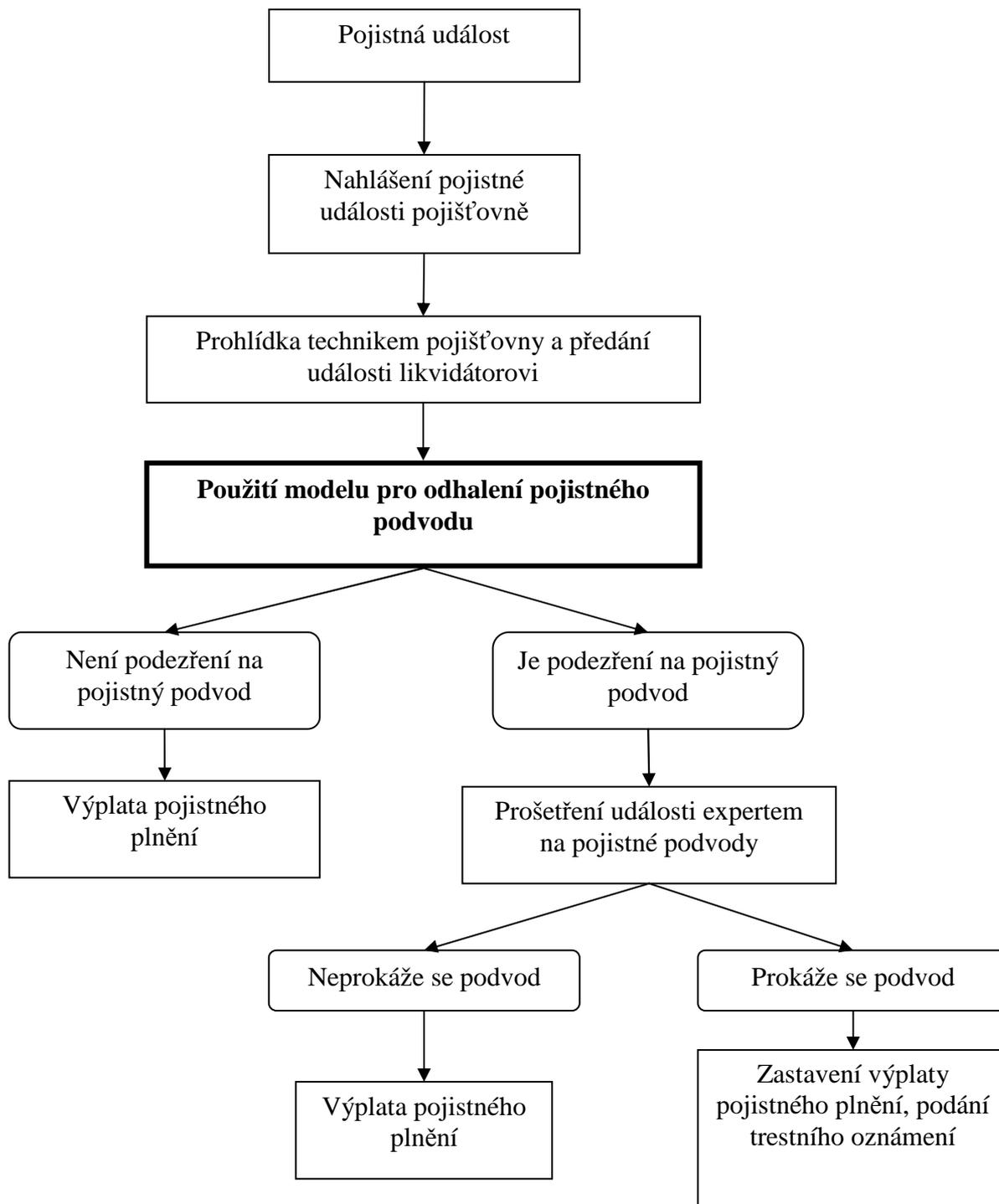
Postup odhalování pojistného podvodu pojištění provozu motorového vozidla by v pojišťovně mohl vypadat jako na obrázku 23.

Nahlášená pojistná událost by se předala likvidátorovi. Ten by s pomocí technika provedl prohlídku poškozeného vozidla. Své záznamy by vložil do softwaru pro odhalení pojistného podvodu. Pokud by se pojistný podvod neprokázal, mohlo by dojít k pojistnému plnění. Pokud by se pojistný podvod prokázal, předal by likvidátor toto šetření specializovanému útvaru na odhalování pojistných podvodů (forenznímu auditu) k došetření. Při prokázání

pojistného podvodu by se celá záležitost předala policii k trestnímu řízení. Pokud by se pojistný podvod neprokázal, proběhla by bez problému výplata pojistného plnění.



Obrázek 22: Model pro detekci pojistných podvodů [zdroj: vlastní]



Obrázek 23: Postup při odhalování pojistných podvodů s použitím modelu pro detekci pojistného podvodu [vlastní [9]]

ZÁVĚR

Tato práce pojednává o využití RM při forenzním auditu v pojišťovně. V první kapitole jsme si přiblížili základní pojmy z pojištění a pojišťovnictví, kterými byly např. pojistník, pojistitel, pojistná smlouva. Také jsme si představili nejběžnější komerční pojištění motorových vozidel. V další části jsme se věnovali rizikům, která mohou nastat jak v pojišťovně, tak i ve sféře podnikatelské, dále pak na nový vědní obor risk management. V následující části byl osvětlen pojem audit a forenzní audit, jako útvar pro odhalování pojistných podvodů. V této kapitole jsme si ukázali několik indikátorů (faktorů) pojistných podvodů, které jsou páchány převážně na pojištění motorových vozidel. V poslední části této práce jsme se pokusili vytvořit model, který by mohl sloužit jako nástroj pro detekci těchto podvodů.

Hlavní cíl této práce byl splněn. Byl představen forenzní audit a osvětleny základy RM pomocí krátkého příkladu ve čtvrté kapitole této práce.

Hlavním přínosem této práce bychom mohli vidět, ve využití modelu pro odhalování pojistných podvodů v pojišťovnách. Tento model by mohl usnadnit práci likvidátorů a specialistů z útvaru forenzního auditu. Také by znamenal ušetření nemalých nákladů pojišťovny na výplatu pojistných plnění.

Páchání pojistných podvodů v souvislosti s pojištěním motorových vozidel se liší od ostatních podvodů svým mezinárodním charakterem, což zvyšuje jeho závažnost. Bylo by třeba zahájit mediální kampaň a přesvědčit veřejnost, že falešná oznámení jsou podvodem a jejich důsledky mohou být velmi vážné. Také by bylo třeba podporovat vývoj nových softwarových produktů, které mohou při detekci značně pomoci.

V dnešní době je aplikována teorie RM do několika nám blízkých odvětví, jakými jsou ekonomika, finanční služby a obchod. Využívá se například pro přidělení úvěrů v bankovníctví atd.

Většího využití by se RM mohlo těšit v pojišťovnictví, kde by se kromě výše zmíněného modelu, dalo využít i pro předpověď pojistitelnosti a nepojistitelnosti rizik, zatřídění rizikových klientů již při uzavírání pojistných smluv, aj. Toho by mohl využít i management, který se zabývá řízením rizik v podniku. Pro včasné odhalení rizikových faktorů, které mohou nastat a jejichž důsledky by mohly být fatální.

V pojišťovnictví se používá mnoho produktů z oblasti umělé a výpočetní inteligence. Právě využití nástrojů na základě RM by mohlo do této oblasti přinést nové možnosti jejich použití.

SEZNAM LITERATURY

- [1] ČAP avizuje za rok 2008 mírný nárůst pojistných podvodů. In Finanční a investiční adresář : Česká republika 2009. Plzeň : Prospector Media, s.r.o., 2009. s. 1-96.
Dostupný z WWW: <prospectormedia.cz>. ISBN 978-80-903233-3-9.
- [2] ČEJKOVÁ, Viktória, et al. Pojišťovnictví. 1. dopl. vyd. Brno : Masarykova univerzita v Brně, 1999. 179 s. ISBN 80-210-1637-X.
- [3] DUCHÁČKOVÁ, Eva. Principy pojištění a pojišťovnictví. 2. aktualiz. vyd. Praha : Ekopress, s. r. o., 2005. 178 s. ISBN 80-86119-92-0.
- [4] Pojišťovnictví : Zákon o pojišťovnictví, Zákon o pojistné smlouvě. 2008 . Ostrava : Sagit, 2008. Dostupný z WWW: <poj.sagit.cz>. ISBN 978-80-7208-687-0.
- [5] Česká asociace pojišťoven [online]. 2007-2008 [cit. 2008-12-20]. Dostupný z WWW: <<http://www.cap.cz/Vyhledavani.aspx?dotaz=Slovn%C3%ADk%20poj%C5%AF>>.
- [6] HRADEC, Milan, KŘIVOHLÁVEK, Václav, ZÁRYBNICKÁ, Jana. Pojištění a pojišťovnictví. 1. vyd. Praha : Vysoká škola finanční a správní, o.p.s., 2005. 216 s. EUPRESS. ISBN 978-80-86754-48-2.
- [7] ČEJKOVÁ, Viktória, MARTINOVIČOVÁ, Dana. Pojišťovnictví. 1. vyd. Brno : Vysoké učení technické v Brně, 2003. 133 s. ISBN 80-214-2404-4.
- [8] Wikipedie, otevřená encyklopedie [online]. 2009 [cit. 2009-03-06]. Dostupný z WWW: <<http://cs.wikipedia.org/wiki/Audit>>.
- [9] SEDLÁČEK, J. Základy auditu. Brno : Masarykova univerzita, 2006. 169 s. ISBN 80-210-4168-4.
- [10] Wikipedie, otevřená encyklopedie [online]. 2009 [cit. 2009-03-06]. Dostupný z WWW: <<http://cs.wikipedia.org/wiki/Forezn%C3%AD>>.
- [11] PORADA, Viktor, PRŠAL, Vlastimil. Vyšetřování trestného činu pojistného podvodu. Pojistné rozpravy. 2008, č. 10, s. 40-61
- [12] GRECO, Salvatore, et al. Rough Sets and Current Trends in Computing. Berlin : Springer-Verlag, 2006. 951 s. ISBN 3-540-47693-8.
- [13] SANKAR, K. Pal, SIMON, C. K. Shiu. Foundations of Soft Case-Based Reasoning. 2nd edition. [s.l.] : A Wiley-Interscience publication, c2004. ISBN 0-471-08635-5..
- [14] PAWLAK, Zdislaw. Rough Sets : Theoretical Aspects of Reasoning about Data. [s.l.] : Kluwer Academic Publisher, 1992. ISBN 9780792314721, s. 3-51.

- [15] VITÓRIA, AIDA. A Framework for Reasoning with Rough Sets. In: PETERS, J. F., SKOWRON, A. Transactions on Rough Sets IV. Berlin : Springer-Verlag, 2005. ISBN 3-540-29830-4, s. 178-276.
- [16] PAWLAK, ZDZISLAW. A Treatise on Rough Sets. In: PETERS, J. F., SKOWRON, A. Transactions on Rough Sets IV. Berlin : Springer-Verlag, 2005. ISBN 3-540-29830-4, s. 1-17.
- [17] POLKOWSKI, L. Rough Sets – Mathematical foundations. Heidelberg: Physica-Verlag company, 2002. ISBN 3-7908-1510-1.
- [18] KOMOROWSKI, Jan, POLKOWSKI, Lech, SKOWRON, Andrzej. Rough Sets: A Tutorial. In: Rough-Fuzzy Hybridization : A New Trend in Decision-Making. Singapore : Springer-Verlag, [1998], s. 3-98.
- [19] RSES [online]. 1994-2005 [cit. 2009-08-08]. Dostupný z WWW: <<http://logic.mimuw.edu.pl/~rses/>>.
- [20] BERKA, Petr. Dobývání znalostí z databází. 1. vyd. Praha : Academia, 2003. 366 s. ISBN 80-200-1062-9.

SEZNAM OBRÁZKŮ

Obrázek 1: Vztah povinného ručení a havarijního pojištění [7].....	13
Obrázek 2: Členění rizik podle velikosti [7]	16
Obrázek 3: Pojistný podvod [11]	22
Obrázek 4: Aproximace množiny [18].....	30
Obrázek 6: RSES2 [19].....	36
Obrázek 7: Metodika CRISP-DM [20]	39
Obrázek 8: upravená data [zdroj: vlastní]	48
Obrázek 9: načtení dat [zdroj: vlastní]	49
Obrázek 10: Informační a rozhodovací systém[zdroj: vlastní]	50
Obrázek 11: split factor [zdroj: vlastní]	51
Obrázek 12: rozdělení dat na trénovací a testovací množinu [zdroj: vlastní].....	51
Obrázek 13: Redukce dat 1 [zdroj: vlastní].....	52
Obrázek 14: Redukovaná množina [zdroj: vlastní].....	53
Obrázek 15: Redukce dat 2 [zdroj: vlastní].....	53
Obrázek 16: Generování pravidel [zdroj: vlastní]	54
Obrázek 17: Generovaná pravidla 1 [zdroj: vlastní]	54
Obrázek 18: Generovaná pravidla 2 [zdroj: vlastní]	55
Obrázek 19: Testování na testovacích datech [zdroj: vlastní]	56
Obrázek 20: Matice záměn [zdroj: vlastní]	57
Obrázek 21: Výsledný model [zdroj: vlastní]	57
Obrázek 22: Model pro detekci pojistných podvodů [zdroj: vlastní]	60
Obrázek 23: Postup při odhalování pojistných podvodů s použitím modelu pro detekci pojistného podvodu [vlastní [9]].....	61

SEZNAM TABULEK

Tabulka 1: Rozdíl mezi interním a externím auditem [vlastní [8]]	20
Tabulka 2: Informační systém [vlastní [13]]	31
Tabulka 3: Rozhodovací systém [vlastní [13]].....	32
Tabulka 4: Seznam atributů[zdroj: vlastní]	40
Tabulka 5: Attr0 [zdroj: vlastní].....	42
Tabulka 6: Attr2 [zdroj: vlastní].....	43
Tabulka 7: Attr3 [zdroj: vlastní].....	43
Tabulka 8: Attr9 [zdroj: vlastní].....	44
Tabulka 9: Attr12 [zdroj: vlastní].....	44
Tabulka 10: Attr13 [zdroj: vlastní]	44
Tabulka 11: Attr14 [zdroj: vlastní].....	45
Tabulka 12: Attr17 [zdroj: vlastní]	45
Tabulka 13: Attr19 [zdroj: vlastní].....	45
Tabulka 14: Attr21 [zdroj: vlastní].....	46
Tabulka 15: Attr22 [zdroj: vlastní].....	46
Tabulka 16: Attr23 [zdroj: vlastní].....	46
Tabulka 17: Attr27 [zdroj: vlastní].....	47
Tabulka 18: Attr28 [zdroj: vlastní].....	47
Tabulka 19: Attr29 [zdroj: vlastní].....	47
Tabulka 20: Attr31 [zdroj: vlastní].....	48

SEZNAM POUŽITÝCH ZKRATEK

IA	interní audit
RM	rough množiny
RSES2	Rough Set Exploration System version 2.2
CRISP-DM	Cross-Industry Standard Process for Data Mining

SEZNAM PŘÍLOH

Příloha 1: CD na zadním přebalu	69
---------------------------------------	----

PŘÍLOHY

Příloha 1: CD na zadním přebalu

OBSAH:

- RealInsuranceFraudDataSet.txt,
- podvod.tab,
- podvod.rses,
- experimenty:
 - o auta.tab,
 - o auta.rses,
 - o nic.tab,
 - o nic.rses,
 - o nic2.rses,
 - o nic3.rses,
 - o nic4.rses,
 - o nic5.rses,
 - o pokus.tab,
 - o pokus.rses,
 - o puffy.tab,
 - o puffy.rses.