

**Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav Systémového inženýrství a informatiky**

**Analýza dat dostupných v sociální politice a jejich využití
při modelování**

Bc. Anna Suková

**Diplomová práce
2012**

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: 2011/2012

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Anna Suková**
Osobní číslo: **E10258**
Studijní program: **N6209 Systémové inženýrství a informatika**
Studijní obor: **Regionální a informační management**
Název tématu: **Analýza dat dostupných v sociální politice a jejich využití při modelování**
Zadávající katedra: **Ústav systémového inženýrství a informatiky**

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


Předpokládá se, že výstupem diplomové práce bude:

- definice základních pojmů ve vybrané oblasti sociální politiky a modelování;
- návrh modelu vybrané oblasti sociální politiky;
- analýza modelu.


Rozsah grafických prací:
Rozsah pracovní zprávy: cca 50 stran
Forma zpracování diplomové práce: tištěná/elektronická

Seznam odborné literatury:

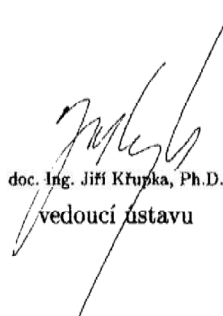
- MELOUN, Milan, MILITKÝ, Jiří. Statistická analýza experimentálních dat. 2.vyd. Praha. Academia, 2004. 953s. ISBN 80-200-1254-0.
DISMAN, Milan. Jak se vyrábí sociologická znalost: příručka pro uživatele. 3.vyd.Praha. Karolinum, 2000. 374s. ISBN 80-246-0139-7.
ŘEHÁK, Jan, ŘEHÁKOVÁ, Blanka. Analýza kategorizovaných dat v sociologii. Praha. Academia, 1986. 397s.
HINDLS, Richard, et.al. Analýza dat v manažerském rozhodování. 1.vyd. Praha. Grada, 1999. 358s. ISBN 80-7169-255-7.
PHILLIPS, David. Quality of life:concept,policy and practice. London. Routledge, 2006. 276s. ISBN 978-0-415-32355-0.
RAPLEY, Mark. Quality of life research: a critical introduction. London. Sage Publications, 2003. 286s. ISBN 978-0-7619-5457-6.

Vedoucí diplomové práce: 
doc. Ing. Jiří Krupka, Ph.D.
Ústav systémového inženýrství a informatiky
Konzultant diplomové práce: **Mgr. Jan Mandys**
Ústav veřejné správy a práva

Datum zadání diplomové práce: **4. října 2011**
Termín odevzdání diplomové práce: **30. dubna 2012**


doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.


doc. Ing. Jiří Krupka, Ph.D.
vedoucí ústavu

V Pardubicích dne 4. října 2011

PROHLÁŠENÍ

Prohlašuji, že jsem tuto práci 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 Univerzity Pardubice.

V Pardubicích dne 20. 4. 2012

Bc. Anna Suková

PODĚKOVÁNÍ:

Tímto bych ráda poděkovala svému vedoucímu práce doc. Ing. Jiřímu Křupkovi Ph.D. za jeho odbornou pomoc, vstřícný přístup, cenné rady a poskytnuté materiály, které mi pomohly při zpracování diplomové práce. Dále bych ráda poděkovala konzultantovi této práce Mgr. Janu Mandysovi Ph.D také za odbornou pomoc, vstřícný přístup, cenné rady a poskytnuté materiály.

ANOTACE

Diplomová práce se zabývá analýzou dostupných dat v sociální politice a jejich následné využití při modelování. Cílem této práce je návrh modelů vybrané oblasti sociální politiky a následná analýza modelů. Pozornost je zde věnována vybraným metodám modelování a definování základních pojmů ve vybrané oblasti sociální politiky. Součástí je několik modelů, kde u každého jsou následně vyhodnoceny získané výstupy.

KLÍČOVÁ SLOVA

Sociální politika, modelování, analýza modelů

TITLE

Analysis of the data available in social policy and their use in modeling

ANNOTATION

This thesis deals with the analysis of data available in the social policy and its subsequent use in modeling. The aim of this work is the model selected areas of social policy model and subsequent models. Attention is paid to selected methods of modeling and defining the basic concepts in selected areas of social policy. There are several models in each case are then evaluated the outputs obtained.

KEYWORDS

Social policy, modeling, analysis models

Obsah

ÚVOD	10
1. DEFINOVÁNÍ VYBRANÝCH POJMŮ Z OBLASTI SOCIÁLNÍ POLITIKY	13
1.1. Sociologie a sociální politika.....	13
1.2. Sociální politika zemí EU	17
1.3. Sociální inženýrství a sociální podnikání	19
2. VYBRANÉ METODY MODELOVÁNÍ	22
2.1. Shluková analýza	22
2.2. Rozhodovací stromy	26
2.3. Asociační pravidla	29
3. VÝBĚR DATOVÉHO SOUBORU	32
3.1. Popis vstupních dat	32
3.2. Analýza vstupních dat.....	33
4. NÁVRH MODELŮ VYBRANÉ OBLASTI SOCIÁLNÍ POLITIKY	40
4.1. Příprava dat a modelování	40
4.2. Model komparace preferencí politický stran voliče v roce 1997 a 2011	41
4.3. Model komparace hodnot a postojů vysokoškoláků v roce 1997 a 2011	51
4.3.1. Model komparace hodnot vysokoškoláků vztahující se k pořízení dětí v roce 1997 a 2011	53
4.3.1. Model komparace hodnot vysokoškoláků vztahující se k uzavření manželství v roce 1997 a 2011	54
4.3.2. Model komparace postojů vysokoškoláků k pořízení dětí v roce 1997 a 2011..	57
4.3.3. Model komparace postojů vysokoškoláků k uzavření manželství v roce 1997 a 2011	62
ZÁVĚR	68
POUŽITÁ LITERATURA	71
SEZNAM PŘÍLOH	74

SEZNAM TABULEK

Tabulka 1: Čtyřpolní kontingenční tabulka	30
Tabulka 2: Datový slovník	37
Tabulka 3: Vývoj preference politických stran voliče od roku 1997 do roku 2011	43
Tabulka 4: Výsledky voleb pravice v roce 1998	44
Tabulka 5: Výsledky voleb levice v roce 1998	45
Tabulka 6: Výsledky voleb v roce 2010	45
Tabulka 7: Srovnání shlukovacích metod	47
Tabulka 8: Čistý měsíční příjem 1997	47
Tabulka 9: Čistý měsíční příjem 2011	48
Tabulka 10: Výstupy z modelovacích technik pro rok 1997	49
Tabulka 11: Výstupy z modelovacích technik pro rok 2011	49
Tabulka 12: Komparace výstupů pro rok 1997 a 2011	50
Tabulka 13: Žebříček hodnot vztahující se k pořízení dětí pro rok 1997 a 2011	53
Tabulka 14: Žebříček hodnot vztahující se k uzavření manželství pro rok 1997 a 2011	54
Tabulka 15: Postoj vysokoškoláků k pořízení dětí v roce 1997 a 2011	58
Tabulka 16: Postoj vysokoškoláků k uzavření manželství v roce 1997 a 2011	63

SEZNAM OBRÁZKŮ

Obrázek 1: Návrh postupu řešení vybraného problému	12
Obrázek 2: Životní kvality a kvalitní život	19
Obrázek 3: Dendogram pro metodu nejbližšího souseda	25
Obrázek 4: Nejčastěji užívané metody shlukování	25
Obrázek 5: Obecný algoritmus pro tvorbu rozhodovacího stromu	27
Obrázek 6: Průběh entropie	28
Obrázek 7: Algoritmus ID3	29
Obrázek 8: Kruhové uspořádání terminálních hodnot podle Schwartze	42
Obrázek 9: Obecný model - typický volič levice/pravice v roce 1997 a 2011	46
Obrázek 10: Obecný model - komparace hodnot vysokoškoláků v roce 1997 a 2011	53
Obrázek 11: Obecný model - komparace postojů vysokoškoláků v roce 1997 a 2011	57
Obrázek 12: Návrh obecného modelu	66
Obrázek 13: Zobrazení nerovnoměrného rozložení dat	80
Obrázek 14: Hodnot mediánu pro nahrazení chybějících dat	80
Obrázek 15: Statistika nahrazených chybějících hodnot	81
Obrázek 16: Postup nahrazování chybějících hodnot	81
Obrázek 17: Vygenerovaná pravidla pro rok 1997 - Apriori – model 1	82
Obrázek 18: Vygenerovaná pravidla pro rok 1997 - metoda GRI	82
Obrázek 19: Výstup z algoritmu TwoStep pro rok 1997- model 1	83
Obrázek 20: Výstup z algoritmu TwoStep pro rok 2011- model 2	84
Obrázek 21: Vygenerovaná pravidla pro rok 1997-Apiority - model 2	84
Obrázek 22: Vygenerovaná pravidla pro rok 2011-Apiority – model 3	85
Obrázek 23: Vygenerovaná pravidla pro rok 2011-GRI - model 3	85
Obrázek 24: Výstup z algoritmu TwoStep pro rok 2011 - model 3	86
Obrázek 25: Výsledný stream z programu Clementine	87
Obrázek 26: Metodika CRISP-DM	88
Obrázek 27: Klasifikace dat	93

SEZNAM GRAFŮ

Graf 1: Preference politické orientace pro rok 1997 a 2011.....	44
Graf 2: Postoj mužů k věku pořízení 1. dítěte	60
Graf 3: Postoj žen k věku pořízení 1. dítěte	60
Graf 4: Zobrazení časové náročnosti jednotlivých etap DM.....	91

Úvod

Sociální politika je téma, které bezpochyby ovlivňuje každého občana, kteréhokoliv státu. V současné době je každodenně diskutovaným tématem ve všech odvětvích problematika úsporných opatření, která se zavádí za účelem snížení deficitu státního rozpočtu. Tato úsporná opatření velmi úzce souvisejí se sociální politikou a to ve všech jejích oblastech. Jedná se například o reformu zdravotnictví, důchodovou reformu či v současnosti velmi diskutovanou reformu školství, která se převážně dotýká vysokého školství. Podstatný problém stálého nárůstu deficitu státního rozpočtu lze spatřovat v tom, že dochází k neustálému růstu některých mandatorních výdajů. Jedná se například o problematiku penzijních důchodů. Výdaje na penzijní důchody ze státního rozpočtu se rok od roku zvyšují a to o nezanedbatelnou výši. V roce 2007 činily tyto výdaje 282,6 miliard korun, v roce 2008 dosahovaly 304,9 miliard korun, v roce 2010 stát vydal na důchodech zhruba o 10% více, než v předcházejícím roce a to téměř 337,5 miliard korun. Podobný stále rostoucí trend se vyskytuje v oblasti zdravotnictví. Stát vydal v roce 2007 do zdravotnictví 206,565 miliard korun, v roce 2009 již výdaje činily 241,637 miliard korun. V roce 2010 došlo opět k nárůstu výdajů, i když v tomto případě k velmi nepatrnému v porovnání s ostatními lety a to na 243,283 miliard korun. I přes stálé zvyšování výdajů ze státního rozpočtu, si celým 48% dotázaných obyvatel České republiky myslí, že výše výdajů vynaložené státem na zabezpečení sociální politiky jsou nízké. Pouhých 35% obyvatel považuje výši výdajů na zabezpečení sociální politiky za přiměřené. Tyto údaje byly získány z průzkumu společnosti CVVM, který proběhl v roce 2010. CVVM představuje centrum pro výzkum veřejného mínění.

Cílem práce je definice vybraných základních pojmů v oblasti sociální politiky, návrh modelů vybrané oblasti sociální politiky a analýza modelů. Pro modelování dat z vybrané oblasti sociální politiky za pomoci data-miningových metod, byla získána data ze dvou zdrojů. Prvním zdrojem byl sociologický ústav AV ČR, v.v.i.¹. Druhý zdroj představoval vlastní dotazníkové šetření. Data byla rozšířena mezi vysokoškolsky vzdělané osoby nebo na vysoké škole stále studující osoby ve věku 18-30 let. Práce je rozčleněna celkem na čtyři kapitoly.

V první kapitole jsou definovány vybrané základní pojmy z oblasti sociální politiky. Definována je zde sociologie, která je považována za jednu ze sociálních věd. Pozornost je také věnována problematice sociální politiky obecně. Je zde vymezen problém, co jsou to

¹ AV ČR, v.v.i. – Akademie věd České republiky, veřejná výzkumná instituce.

objekty a subjekty sociální politiky, jaké jsou jednotlivé funkce sociální politiky a základní principy a znaky sociální politiky. Dále je zde definována problematika týkající se sociální politiky zemí Evropské unie. V současné době, jsou velmi často diskutované dva termíny týkající se sociální politiky a to sociální inženýrství a sociální podnikání. Proto je i určitá část této kapitoly věnována těmto dvěma problémům.

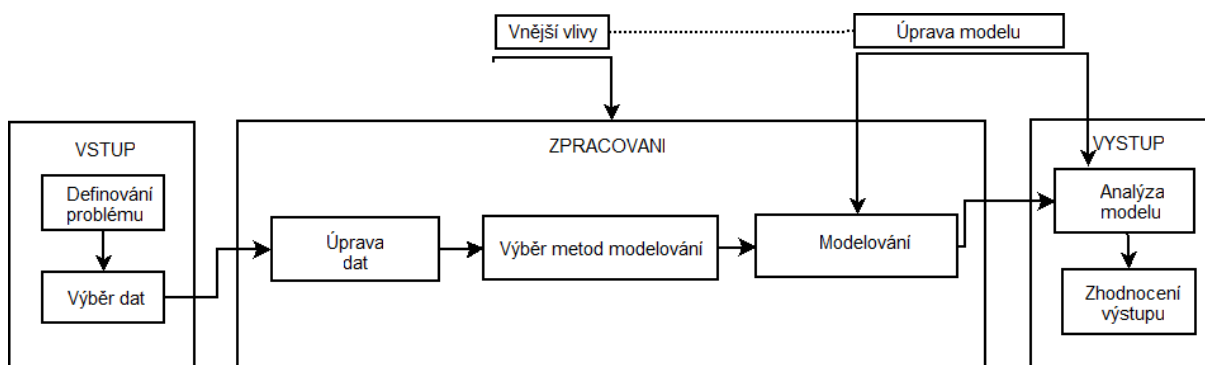
V druhé kapitole se zabývám jednotlivými vybranými modelovacími metodami, které lze využít pro modelování ve vybrané oblasti sociální politiky a které také byly v této práci aplikovány na data. Jedná se konkrétně o asociační pravidla, rozhodovací stromy a shlukovou analýzu.

Následující třetí část práce je věnována výběru vstupních dat. Je zde uveden popis obou zdrojů, ze kterých vstupní data pocházejí. Vstupní data, která se označují jako atributy, jsou zde vyjmenována a zkráceně charakterizována. Součástí této kapitoly je také datový slovník. Po načtení dat do programového prostředí byla provedena základní statistická analýza dat, která je uvedena v příloze B.

Poslední kapitola této práce je věnována přípravě data a návrhu modelů. Začátek této kapitoly se věnuje přípravě dat a posouzení kvality dat. Data byla zkontrolována, zda neobsahují chybějící či chybné záznamy. V případě zjištění určitých problémů, byla data v programovém prostředí upravena. Další část této kapitoly je stěžejní. Je věnována analýze dat, návrhu modelů a následnému zhodnocení modelů a vyvození konkrétních závěrů. V této kapitole jsou nejprve analyzována data, dále jsou popsány postupy tvorby jednotlivých modelů, jejich různé nastavení a jsou provedeny analýzy na základě výsledných modelů. Grafické výstupy jednotlivých modelů jsou převážně uvedeny v přílohách. V příloze G je také umístěn výsledný stream získaný z programového prostředí Clementine. V závěru této kapitoly je krátce rozebrána metodologie CRISP-DM, jednotlivé typy dat, tak jak na ně data-mining nahlíží. Krátce je zde zmíněno o programovém prostředí Clementine, ve kterém bylo realizováno modelování s daty a také o typech dat, se kterými lze pracovat v programu Clementine. Podrobnější definování výše zmíněné problematiky je uvedeno v příloze G.

Celkový postup práce na této diplomové práci je graficky vyjádřen na obrázku 1. Jsou zde zahrnuty jednotlivé postupy práce, které jsou součástí všech uvedených kapitol, krom první. Z obrázku jsou jasně patrné vstupy do systému, které tvoří prvotní definování problému a následný výběr dat. Součástí bloku zpracování je úprava dat, výběr vhodných modelovacích metod a modelování. Výstup ze systému tvoří popis modelů a následné zhodnocení výsledků. Součástí modelu je nezbytná zpětná vazba, úprava modelů. Na systém má ve většině případů

vliv působení vnějších vlivů, proto i zde je tento fakt zohledněn. Mezi vnějšími vlivy, které většinou tvoří nepřesnosti a různé typy chyb a úpravou modelů existuje jistý vztah. Pokud se vyskytne jistá nepřesnost nebo chyba, je nutné provést jistou úpravu modelů a tím odstranit dané problémy.



Obrázek 1: Návrh postupu řešení vybraného problému

Zdroj: upraveno podle [17]

1. Definování vybraných pojmů z oblasti sociální politiky

V této první části práce se budu věnovat základním pojmům z oblasti sociologie a sociální politiky. Podrobněji tu bude rozebrána problematika týkající se sociologie a problematika sociální politiky obecně. Určitá část je zde věnována i konkrétní problematice sociální politiky v zemích EU a s ní související poměrně důležitá Evropská sociální charta. Závěr této první části je věnována dvou důležitým termínům z oblasti sociální politiky, které se v současné době velmi často objevují v diskuzích. Jedná se o problematiku sociálního inženýrství a sociálního podnikání. S problematikou sociálního inženýrství je velmi často spojován současný prezident ČR prof. Ing. Václav Klaus, CSc., proto je v této práci podrobněji také rozebrán jeho názor na danou problematiku.

1.1. Sociologie a sociální politika

Sociologie je považována za jednu ze sociálních věd. Nejedná se o normativní vědu, to znamená, že nám neříká, jaká má společnost být, ale zkoumá společnost, jaká ve skutečnosti je. Jedná se o vědu, která hledá, zkoumá a pochybuje o všem a to včetně sebe samé. Sociologie používá obecnou metodologii vědy a má svůj předmět. Předmětem sociologie jsou sociální jevy. Můžeme vlastně říci, že vše co se děje mezi lidmi, je sociální. To, že jsou lidé spolu a sdružují, vytváří sociální děje. Velmi často se můžeme také setkat s označením sociologie, jako věda o společnosti. A to z toho důvodu, že v podstatě tvoří společnost. Sociologie se zabývá tedy tím, co se děje mezi lidmi a nikoliv tím, co se děje v lidech. [14]

Jak již bylo řečeno výše, za předmět sociologie jsou považovány sociální jevy. Popis těchto sociálních jevů se označuje jako tzv. sociografie. Jedná se o nejjednodušší první krok zkoumání. Poté následuje analýza, která hledá a zkoumá vzájemné souvislosti sociálních jevů a měří intenzitu zjištěných závislostí. V sociologii se nepracuje s různými zákony. A to v důsledku toho, že po desetiletích zkoumání dospěli k tomu, že sociální dění není nijak zákonité. Zamítli tedy problematiku příčiny a důsledku a přijali tzv. princip pravděpodobnosti.

Za zakladatele sociologie je považován August Comte, který v roce 1839 poprvé uvedl pojem sociologie ve svém díle „Cours de Philosophie Positive“. Výrazná sociální změna, která způsobila Velkou francouzskou revoluci, zapříčinila i vznik sociologie. Obecně sociologie tedy nevznikla z toho důvodu, že existovaly sociální jevy, ale z toho důvodu, že sociální jevy postihly poměrně rychlé změny. [14]

Sociální politika

Pro definování pojmu sociální politika se lze setkat s velmi rozmanitým vymezením. V Ottově naučném slovníku, je problematika sociální politiky definována takto: „*Sociální politika je především praktická snaha, aby společenský celek byl uspořádán co nejideálněji. Jako nauka jeví se nám pak sociální politika jako souhrn zásad, jak odstraniti či alespoň zmírniti vady společenského života.*“ [13, str. 62]

V jiné literatuře se lze setkat s vymezením pojmu sociální politika jako praktické úsilí, které má vést k tomu, aby společenský celek byl vychován a vytvořen co nejideálněji. Hybnou pákou sociální politiky není milosrdenství, ale spravedlnost a společenská účelnost. Tento názor na problematiku pochází od prof. Engliš z roku 1916. [38]

Definování problému sociální politiky, lze naleznou i v Masarykově slovníku naučném. Zde ale spíše názor na sociální politiku souvisí s chápáním samotného pojmu „sociální“. Termín „sociální“, lze chápat ze třech různých pohledů. Na pojem sociální můžeme nahlížet v nejširším slova smyslu, tj. sociální jako společenský. V užším slova smyslu, chápeme sociální jako snahu směřující k vyrovnání nerovností a vážící se bezprostředně k životním podmínkám lidí. V nejužším slova smyslu chápeme sociální z hlediska řešení nepříznivých nebo nouzových sociálních situací.

Pro zodpovězení otázky, co je sociální politika, je dobré vycházet z obecného pojmu politiky vůbec. Pod pojmem politika můžeme chápat specifickou společenskou činnost, (typické jsou pro ni soubory různých opatření), konkrétní jednání různých subjektů na odlišných úrovních (ne jenom státu), jimiž je ovlivňována společenská realita v nejširším slova smyslu. Tato specifická společenská činnost slouží k aplikování stanovených cílů, které jsou stanovovány především na základě vědeckého poznání. Od politiky se tedy očekává, že působí ve smyslu objektivních trendů, které jsou vědecky podloženy. To tedy znamená, že jako činnost předpokládá vysokou odbornou úroveň a profesionalitu, včetně toho, umět správně hodnotit podmínky, čas, prostředky, způsoby, formy a metody tvorby, posuzování a realizování cílů. Toto obecné definování pojmu politiky lze i použít na pojem sociální politiku s tím, že má určitý vliv nikoli na společenskou, ale sociální realitu. Sociální reality bezpochyby souvisí s vymezením sociální sféry. Problém ale je, že sociální sféru nelze vymezit odvětvově. Nahlížíme na ni jako na sféru samostatnou, průřezovou, za systém sociálních vztahů, podmínek a procesů, které se projevují v obsahu a charakteru činností, v chování, v potřebách, v zájmech, hodnotách a cílech sociálních subjektů. To, že ji nejsme schopni nijak explicitně vymezit, neznamena ale, že sociální sféra reálně neexistuje. Znamená

to však, že jako referenční rámec sociální politiky je značně neurčitá. Z tohoto důvodu se také v současnosti setkáváme s odlišným chápáním a vymezením pojmu sociální teorie. [2,38]

K subjektům sociální politiky patří zejména stát a jeho orgány. Důležitá role zde připadá parlamentu. Na základě zvoleného programu a jemu příslušné koncepce vytváří a schvaluje legislativní opatření, která mají velmi důležitý význam pro sociální politiku. V tomto případě se nemyslí pouze státní sociální politika. Za subjekt sociální politiky můžeme dále považovat i zaměstnavatele. Ti jsou povinni zajistit opatření, která nařizuje stát a jeho orgány. Dále mezi subjekty sociální politiky lze považovat zaměstnavatelské, zaměstnanecké odborové orgány, dále občanské iniciativy, církve, či občany, rodiny a domácnosti. [2,38]

Při realizaci sociální politiky je nutné dodržovat určité základní principy. To znamená základní zásady a pravidla činnosti, hlavní myšlenky a myšlenkové postupy, které jsou pro sociální politiky nezbytné a určující. Jedná se tedy o principy, které spolu velmi úzce souvisí a které jsou pro sociální politiku nejdůležitější. Rozlišujeme tedy:

- princip sociální spravedlnosti;
- princip sociální solidarity;
- princip subsidiarity;
- princip participace.

Princip sociální spravedlnosti je klíčovým principem sociální politiky. Jedná se o jakési měřítko mezilidských vztahů, za základní mravní a náboženskou hodnotu či za základ právního řádu, států a státních politik. Princip sociální spravedlnosti představuje jednak spravedlnost v právním slova smyslu, ale také se jedná o spravedlnost sociální. Tuto sociální spravedlnost lze vymezit pravidly, kterými jsou ve společnosti rozlišovány předpoklady a prostředky veřejného blahobytu mezi jednotlivé subjekty. Vzhledem k sociální politice jsou důležitá obě vymezení. Obecně solidaritu chápeme jako jakousi vzájemnou podporu či sounáležitost. Princip sociální solidarita je spojena hlavně s vznikem a selektováním životních podmínek a prostředků jedinců a sociálních skupin v zájmu naplňování idee sociální spravedlnosti. Princip subsidiarity je v oblasti sociální politiky stále čím dál tím více diskutovanější. Podstata principu vychází z pohledu na člověka jako na individuuum s jedinečnými vlastnostmi, schopnostmi, dispozicemi, možnostmi atd., které jedinec „musí“ využívat ku prospěchu a sociálnímu bezpečí svému i druhých. Princip subsidiarity nám říká, že každý jedinec je povinen nejdříve pomoci sám sobě a v případě, že nemá tuto možnost, je povinna mu pomoci rodina. Rodina je v takovém případě povinna nejdříve pomoci si sama,

svými vlastními silami a prostředky a až teprve poté si na pomoc vyžádat jiná společenství. Těmito společnostmi se myslí např. přátele, obec aj. Podobně se v takovém případě zachová i obec a až na posledním místě je na pomoc přizván stát. Stát má povinnost podporovat a sám vytvářen podmínky, aby si každý mohl pomoci sám a stát sám nemusel zasahovat, popř. aby zasahoval na posledním místě. Princip participace spočívá vlastně v tom, aby lidé měli možnost spolu účastnit se na všem, co se dotýká jejich vlastního života (zdraví, zajištění v nemoci, ve stáří, aj.). Praxe nám hovoří o tom, že pokud lidé nemají možnost nebo se neúčastní z jiných důvodů takovýchto sociálně politických opatření, jsou výsledky opatření často méně efektivní, než je žádáno a často jsou i sebelepší záměry zbytečné. [38]

Schopnost vyvolávat určité efekty, schopnost sledovat a realizovat určitou vědomou činnost a úkoly a schopnost vyvolávat účinky ve vztahu k objektům označujeme za funkci sociální politiky. Rozlišujeme několik různých funkcí sociální politiky. Pro jejich klasifikaci rozlišujeme různá kritéria. Jedná se o tyto funkce:

- ochranná;
- rozdělovací a přerozdělovací;
- homogenizační;
- stimulační;
- preventivní.

Kromě těchto výše uvedených funkcí můžeme také z jiného pohledu rozlišovat funkci zabezpečovací, vyhledávací, sociálně diagnostické, kontrolní aj., které plynou z Listiny práv a svobod občana, přijaté FS ČSFR². [38]

Ochranná funkce představuje řešení již existujících sociálních událostí, kdy jedinec nebo sociální skupina se dostala do určité nevýhody vzhledem k ostatním jedincům či ostatní sociální skupině. V tomto případě se jedná jak o ekonomické znevýhodnění, tak znevýhodnění v sociálním smyslu. Jedná se o funkci nejstarší, ze všech výše uvedených. Ochranná funkce tvoří tradiční a stabilní prvek sociální politiky, který je nutné neustále podporovat a posilovat. Rozdělovací a přerozdělovací funkce je považována za jednu z nejsložitějších, ale také za nejvýznamnějších funkcí. V určitém pohledu lze říci, že touto funkcí je prováděn vlastní obsah opatření sociální politiky. Hlavní směr, který sleduje je problematika příjmů, mezd a důchodů a jejich redistribuci. Dále této funkci náleží i parametr rozdělování životních šancí, situací, které vznikají díky sociální politice. Podstatou přerozdělovací funkce je v přeměně prvotního rozdělování v přerozdělení toho, co již jednou

² Federální shromáždění České a Slovenské federativní republiky.

bylo nedokonale či „nespravedlivě“ rozděleno trhem. Za poměrně novou funkci sociální politiky je považována funkce homogenizační. Poměrně úzce je spjata s funkcí rozdělovací a přerozdělovací. Podstatou této funkce je zvýšení stejnorodosti v uspokojování základních životních potřeb a zvýšení stejnorodosti zájmů a hodnot. Úkolem stimulační funkce je podporovat, podněcovat, vyvolávat žádoucí sociální jednání jednotlivců a sociálních skupin a to v oblasti ekonomické i ostatní. Velmi často se také lze setkat s označením této funkce jako funkce produktivity. Poslední funkcí sociální politiky, která je uvedena výše, je preventivní funkce, někdy také označovaná jako profylaktická. Souvisí se snahou předcházet určitým ztrátám na životě a zdraví, se snahou předcházet nepříznivým sociálním situacím, jako například chudoba či nezaměstnanost. Preventivní funkce v posledních letech velmi silně roste. Je to způsobeno zejména tím, že moderní sociální politika se neomezuje pouze na ochranu a pomoc v situacích nouze, ale znamená širokou existenční prevenci zaměřenou na zvládání životního stylu. [2,38]

1.2.Sociální politika zemí EU

Důležitý moment v oblasti sociální politiky vznikl pro Českou republiku dne 17.1.1996, kdy podala ČR oficiálně přihlášku o členství do Evropské unie. Pro přijetí do EU bylo nutné, aby kandidátská země splnila určité podmínky v oblasti politické, ekonomické, ale také v oblasti sociální politiky. Bylo tedy nezbytné studovat kromě jiného zkušenosti a přístupy k sociální politice jednotlivých zemí EU. Kromě jiného bylo také nutné, aby se příslušné orgány ČR seznámili s tendencemi a závěry, které se za celý komplex zemí EU prosazují v této oblasti. Dále bylo nezbytné se věnovat tomu, jak flexibilně zavádět jednotlivé direktivy a regulace EU na úseku sociální politiky do praxe ČR.

K cílům sociální politiky ve vyspělých zemích a zemích s dobře fungující tržní ekonomikou patří:

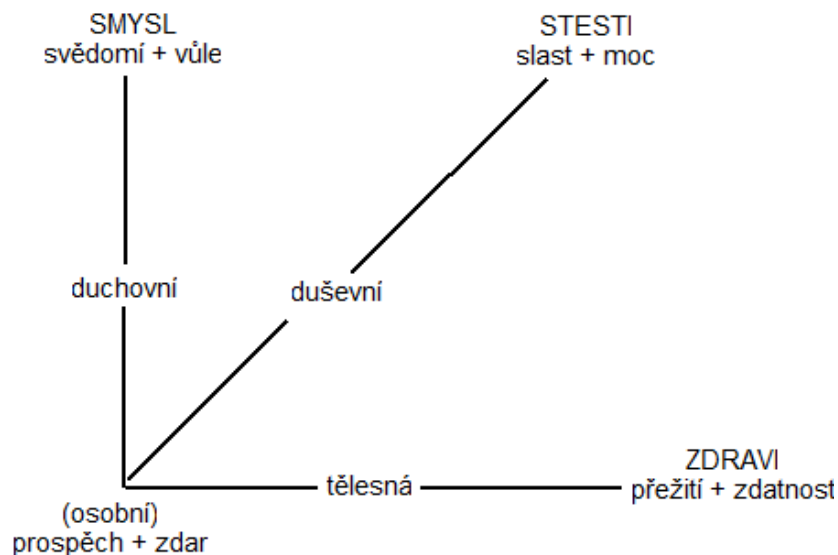
- *„přispívat ke zvyšování výkonnosti ekonomiky (prostřednictvím zvyšování schopností lidí a jejich uplatnění;*
- *Přispívat k humanizaci společnosti (uspokojování sociálních potřeb, resp. Vytváření podmínek pro toto uplatňování tak, aby byl zabezpečen aktivní a důstojný život členů společnosti).“ [2, str.9-10]*

Sociální politika v zemích EU, ve kterých lze považovat tržní ekonomiku za poměrně úspěšnou je z několika důvodů velmi úzce spojena s hospodářskou politikou. Obě tyto

politiky se velmi úzce ovlivňují. Aby byla účinná sociální politika, je nezbytná výkonná ekonomika, neboť na níž závisí tvorba prostředků pro sociální politiku. Na druhou stranu výsledky hospodářské politiky závisí na tom, jak jsou aktivní komponenty sociální politiky. Tedy na tom, jak jsou připraveni a motivováni lidé k odpovědné činnosti.

Sociální politika EU v současnosti se snaží díky sociálnímu právu a sociálním institucím o sociální ochranu lidí a tím o ochranu integrity jejich osobnosti. Evropské země prostřednictvím soustavy mezinárodních ujednání usilují o vytvoření minimálních norem pro zabezpečení základních lidských práv. Pokud porovnáme podporu pro nízko příjmové či neúplné rodiny v zemích EU a ve vyspělé USA, země EU poskytují více podpory než USA. Výsledky tohoto tvrzení je vyvozeno ze srovnání sociálních programů obou zájmových oblastí. Téměř všechny evropské země mají zakotveny obsáhlé programy sociální politiky, které vedou k vyšší podpoře než je tomu tak v USA. Cílem EU je zajistit ve všech členských zemích minimální sociální standard. Na jednotlivých zemích se již ponechává úsilí a možnosti, jak se dostat nad tuto minimální hranici. [2]

Důležité je, se ještě zmínit o sociální kvalitě versus kvalita života. O pojmu kvalita života lze říci, že je nutně stejně multidimenzionální jako je lidský život. Obecné rozlišení dimenzí, v nichž pojmově a operacionálně kvalitu života vymezujeme a zjišťujeme, v zaměření člověka vůči jeho světu a vůči sobě samému ukazuje obrázek 2, kde jsou uvedeny některé typické znaky pozitivního zacílení v nich. Sociální kvalita má být komplexní a zahrnovat jak objektivní, tak i subjektivní interpretace. Existuje mnoho podobností mezi přístupem ke kvalitě života a k sociální kvalitě. Snad nejdůležitější podobností však je, že oba pocházejí ze stejných evropských tradic, které jsou založeny na základních cílech EU v oblasti posilování sociální soudržnosti a v boji proti sociálnímu vyloučení. V této souvislosti jsou oba zjevně normativní ve svých přístupech k sociální soudržnosti. Oba mají silné sociální politiky zaměřující se na naplňování sociální spravedlnosti prostřednictvím celoevropské sociální politiky. Jejich geneze je však poněkud odlišná. Zastřešující kvalita života je pevně zakotven v rámci rubriky EU, kde jsou cíle stanovené smlouvami a zákony, zatímco sociální kvalita vychází z kritiky a protiváhy toho, že v rámci EU je spíše kladen důraz na ekonomickou koncepci než na sociální koncepci kvality života. Na sociální kvalitu lze nahlížet jako na „humanistické protilátka“ k vysoce utilitární a všudypřítomná praxi, týkající se srovnávání kvality života s příjmem nebo majetkem. Takže jedním hlavním impulsem pro výstavbu sociálního kvality je výroba alternativních opatření k HDP pro hodnocení kvality společností. [18]



Obrázek 2: Životní kvality a kvalitní život

Zdroj:[22]

Evropská sociální charta

Evropská charta byla podepsána 18.10.1961 v Turinu a nabyla platnosti 26.2.1965 po ratifikaci pěti států. Představuje absolutní předěl na cestě sjednocování sociální politiky západoevropských zemí. Protokol, který novelizoval Evropskou sociální chartu, byl přijat v Turinu v říjnu 1991. „*Evropská sociální charta zakotvuje ochranu sociálních hospodářských práv v širokém rozsahu.*“ [2, str.32]. Nejedná se o právně závazný dokument, ale jde především o slavnostní vyhlášení základních principů, z nichž má vycházet evropské chápání sociální politiky. Tvoří právní rámec pro budoucí postupnou harmonizaci sociální politiky v zemích Evropského společenství. Sociální charta představuje tedy významný bod v dlouhodobé snaze prosazování stejné důležitosti sociální politiky jako aspektům ekonomickým. [2]

1.3.Sociální inženýrství a sociální podnikání

Poměrně často diskutovaným termínem ve společnosti je termín „sociální inženýrství“. Často je s tímto termínem spojován současný prezident prof. Ing.Václav Klaus, CSc. Zdá se, že v samém jádru Klausova uvažování je jeho silné přesvědčení o přirozenosti volného trhu, o racionalitě aktérů a přirozenosti moderního národního státu. Všechny tyto tři fenomény se v Klausově výkladu stávají neoddělitelnou součástí výbavy lidské společnosti, přičemž různé „síly“ se snaží tuto přirozenost narušovat. Velmi často se ve výkladech Václava Klause

můžeme setkat s označením „sociální inženýři“. Pojmem „sociální inženýři“ Klaus označuje své oponenty. Označuje je tak vlastně za „ničitele“ přirozených pořádků. Dramatické proměny posuzuje podle toho, zda odpovídají jeho definici přirozeného. Pokud změny zahrnují například posílení volného trhu, pak jde o návrat k přirozenému stavu, takový zásah pak nemůže být kritizovatelný, přirozenost je totiž stavem nejvhodnějším. Bez ohledu na to, že i taková transformace má dramatické dopady na celou společnost, Václav Klaus by je za sociální inženýrství neoznačil. Pokud se však změny odehrávají odlišným způsobem než tím, který Klaus upřednostňuje, pak jde podle Klause o „sociální inženýrství“ (nadmárodní celky, regulace trhu atd.). Přitom vlastně každá transformace (i ta Klausova z počátku 90. let) má vliv na životy lidí a jako taková je určitým sociálním inženýrstvím. Václav Klaus jakoby ve své rétorice zapomínal či zamlčoval, že každý zápas o uspořádání společnosti je v posledku také zápasem ideologickým, nikoliv bojem za konečné prosazení přirozeného stavu. [6]

Sociální podnikání

Sociální podnikání se od obecně známého termínu „podnikání“ především liší tím, že jeho hlavním cílem není pouze zisk. Snahou je také opatřit volné prostředky, které majitel firmy využije na veřejně prospěšnou věc. Velmi často se stává, že zaměstnavatel již sociálně podniká, jen třeba netuší, že pro to existuje nějaký termín. Jeden z prvních zastánců tohoto konceptu v moderní době je Muhammad Yunus³, jehož snahou je snižovat chudobu v rozvojových zemích prostřednictvím poskytováním výhodných mikropůjček chudým živnostníkům. V roce 2006 získal společně s bankou Grameen Bank Nobelovu cenu za mír za své „úsilí vytvářet hospodářský a sociální rozvoj odspodu“. V ČR chybí legislativní úprava pro podnikatele, kteří se snaží podnikat podobným způsobem, který je prospěšný pro společnost. Je velmi časté, že sociálně prospěšný podnikatel často zaměstnává sociálně nebo zdravotně znevýhodněné občany.

Na otázku, ***jak bude vypadat budoucnost sociálního podnikání***, odpovídá zkráceně Christian Seelo⁴ v článku „Peníze nejsou cestou ke snižování chudoby“ takto: „*Z desetiletí tradičních snad o rozvoj, kdy se vyplácely každoročně miliardy dolarů, víme, že peníze nejsou cestou ke snižování chudoby. Skutečným řešením je vymyslet chytřejší projekty, jež by využily zdroje výnosněji a z malého ekonomického vstupu by vytvořily sociální hodnotu. Obvyklé vývojové postupy příliš často selhávají a sociálně prospěšné podnikání se stane stále*

³ Zakladatel bangladéšské Grameen Bank a vlastník Nobelovy ceny za mír

⁴ Ředitel IESE platformy pro strategii a udržitelný rozvoj

důležitějším fenoménem k rozpoznání a schválení nových trajektorií vývoje. Z mého náhledu vyplývá, že to může být nejen plodná cesta vytvářející budoucnost chudých zemí, ale také přehodnocení procesů v zemích OECD⁵ například prostřednictvím návrhu organizačních modelů, které jsou soutěžní, avšak zároveň maximalizují pracovní úsilí.“ [27, str. 3]

⁵ Organizace pro hospodářskou spolupráci a rozvoj

2. Vybrané metody modelování

Tato kapitola se zabývá popisem a analýzou jednotlivých možných modelovacích metod, které lze využít pro analýzu dat. Jednotlivé metody je nutné vybírat s ohledem na typ dat, na požadované výsledky aj. Mezi vybrané metody pro analýzu dat byla zvolena shluková analýza (seskupovací, clusterová), rozhodovací stromy a asociační pravidla.

2.1. Shluková analýza

Shluková analýza je řazena mezi metody, které se zabývají vyšetřováním podobnosti vícerozměrných objektů a jejich klasifikací do tříd neboli shluků. Jedná se o objekty, které obsahují větší množství proměnných. Shluková analýza se tedy zabývá řešením otázky, zda lze pozorované příklady rozdělit do shluků (skupin) vzájemně si blízkých příkladů. Principem je, že všechny objekty vně shluků jsou si co nejvíce podobné a naopak objekty různých shluků jsou navzájem co nejvíce nepodobné. Předpokládá se, že lze měřit vzdálenosti mezi příklady. V shlukové analýze je ale neznámá příslušnost do tříd všech objektů. Neznámý je i počet tříd (shluků). U shlukové analýzy lze rozlišit tři hlavní cíle:

- *„Popis systematiky- jedná se o klasické využití shlukové analýzy pro průzkumové cíle a taxonomii (empirická klasifikace objektů);*
- *Zjednodušení dat- při hledání taxonomie nám analýza shluků nabízí zjednodušený pohled na objekty;*
- *Identifikaci vztahu- po nalezení shluků objektů, a tím i struktury mezi objekty, je jednodušší zobrazit vztahy mezi objekty.“ [12, str. 454-455]*

Tyto cíle shlukové analýzy není možné oddělit od hledání a volby vhodných znaků k popisu seskupovaných objektů. Volení nejvhodnějších znaků je nutné provádět na základě podkladu teoretických, pojmových a praktických skutečností. Součástí shlukové analýzy není rozlišování významných a nevýznamných znaků. Jedná se pouze o odlišení jednotlivých shluků. Rušivý efekt na výsledky shlukové analýzy může být způsoben nesprávným zařazením znaků, což představuje zahrnutí i odlehlých objektů. Proto je vhodné vždy pracovat pouze s takovými znaky, které dostatečně rozlišují mezi objekty. [1, 12]

Nejčastěji užívané míry, založené na prezentaci objektů v prostoru, jehož souřadnice tvoří jednotlivé znaky, označujeme jako *míry vzdáleností*. V případě, že tyto míry splňují podmínky symetrie

$$d(x, y) = d(y, x) \quad (1)$$

a trojúhelníkovou nerovnost

$$d(x, y) \leq d(x, z) + d(y, z) \quad (2)$$

jedná se o tak zvané *metriky*. Nejčastěji používanou vzdáleností mírou je *euklidovská vzdálenost*. Výpočet euklidovské vzdálenosti je založen na Pythagorově větě. Představuje délku přepony pravoúhlého trojúhelníku.

$$d_e(x_k, x_l) = \sqrt{\sum_{j=1}^m (x_{kj} - x_{lj})^2} \quad (3)$$

Mezi další velmi často používané metriky řadíme i *manhattanskou vzdálenost*, někdy také označovanou jako *vzdálenost městských bloků* nebo *Hammingovu metriku*.

$$d_H(x_k, x_l) = \sum_{j=1}^m |x_{kj} - x_{lj}| \quad (4)$$

Před použitím manhattanské metriky je nutné, aby znaky spolu nekoleovaly. V případě, že toto není splněno, shluky jsou nesprávné. Dále rozlišujeme několik dalších využívaných metrik, jako například Čebyševovu metriku.

Výše uvedené míry vzdáleností závisí na měřítku veličin. Je tedy nutné dané veličiny normovat. Normování se provádí dělením konkrétní hodnoty nějakou jinou hodnotou. Jedná se například o průměr, směrodatnou odchylku nebo rozpětím (max-min). Dále je třeba, aby u všech veličin byl stejný rozptyl. [8, 12]

Metody shlukové analýzy lze rozlišit podle systému použité klasifikace na:

- Hierarchické metody
- Nehierarchické metody

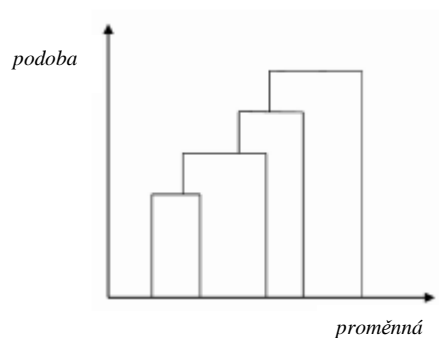
Pro hierarchické metody je typické, že každý shluk je současně podmnožinou jiného shluku s výjimkou samotné množiny objektů, která je považována za maximálně možný shluk. Hierarchické shlukovací postupy vycházejí z hierarchického uspořádání objektů a jejich shluků. Grafickým výstupem hierarchicky uspořádaných shluků jsou dvě formy a to buď vývojový strom nebo dendrogram. U hierarchických metody ještě rozlišujeme dvě základní skupiny:

- Aglomerativní
- Divizní

Tyto dvě základní skupiny se od sebe liší způsobem shlukování. U aglomerativního přístupu se vychází od jednotlivých objektů a ty se postupně shlukují do hierarchického systému podmnožin, až dokud není dosaženo konečné spojení všech objektů do množiny objektů O . Naopak u divizního přístupu vycházíme z množiny objektů pro klasifikaci jako celku a dále se postupně jednotlivé objekty rozdělují. Takto získáváme hierarchický systém podmnožin. Mezi metody metriky shlukování řadíme:

- Metoda nejbližšího souseda
- Metoda nejvzdálenějšího souseda
- Metoda průměrné vzdálenosti
- Wardova metoda
- Metoda těžiště
- Metoda mediánová

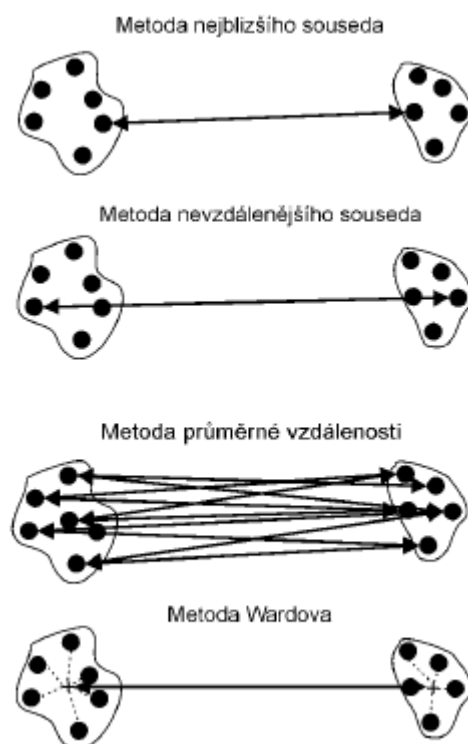
Postup u metody nejbližšího souseda vychází z minimální vzdálenosti. Jsou vždy nalezeny dva objekty, jejichž vzdálenost je nejkratší a umístí se do shluku. Další shluk je vytvořen přidáním třetího nejbližšího objektu. Proces shlukování je ukončen v okamžiku, kdy jsou všechny objekty v jednom společném shluku. Vzdálenost mezi dvěma shluky lze tedy definovat jako nejkratší vzdálenost libovolného bodu ve shluku vůči libovolnému bodu ve shluku jiném. Dva shluky jsou tedy spojeny v libovolném stadiu nejkratší spojkou. Nevýhodou metody nejbližšího souseda, která je velmi častá, je řetězový efekt. Jedná se o situaci, kdy se spojují shluky, jejichž dva objekty jsou sice nejbližší, ale vzhledem k většině ostatních objektů nejde o nejbližší shluky. Klasický dendrogram pro metodu nejbližšího souseda je uveden na obrázku 3, kde na ose x je uvedena proměnná (variables) a na ose y podoba (similarity) [12]



Obrázek 3: Dendogram pro metodu nejbližšího souseda

Zdroj:[12]

Metoda nejvzdálenějšího souseda je velmi podobná předešlé metody, s tím rozdílem, že kritérium je postavena na maximální vzdálenosti, nikoliv tedy na minimální. Tato metoda je také často nazývána metodou úplného propojení. Kritériem u metody průměrné vzdálenosti je logicky průměrná vzdálenost všech objektů v jednom shluku ke všem objektů ve druhém shluku. U metody těžiště (Wardova metoda) je vzdálenost dvou těžišť shluků, která je vyjádřena Eukleidovskou vzdáleností nebo čtvercem Eukleidovské vzdálenosti. Grafické zobrazení jednotlivých metod zobrazuje obrázek 4. [12]



Obrázek 4: Nejčastěji užívané metody shlukování

Zdroj:[8,12]

Nehierarchické metody členíme na:

- Optimalizační metody
- Metody analýzy modů [8]

2.2. Rozhodovací stromy

Rozhodovací stromy řadíme mezi analytické nástroje, které využíváme k nalézání pravidel a vztahů v datovém souboru prostřednictvím postupného rozdělování a větvení na nižší úrovně. Indukce rozhodovacích stromů se řadí mezi nejznámější algoritmy z oblasti symbolických metod strojového učení. Při vytváření rozhodovacího stromu se postupuje podle metody „*rozděl a panuj*“. Při tvorbě rozhodovacího stromu pracujeme s trénovacími daty, které na počátku představují jednu množinu. Tato množina je postupně rozdělována na menší a menší podmnožiny tak, aby v těchto podmnožinách převládaly příklady jedné třídy. Výsledkem této operace jsou tedy podmnožiny tvořené příklady téže třídy. Tento postup se nazývá „top down induction of decision trees“ (TDIDT). Problémem může být určení, na kolik „větví“ se má dělit každá proměnná. V případě, že zvolíme příliš mnoho proměnných, může nastat problém, kdy do každé z těchto skupin přísluší pouze několik málo záznamů a nelze tak vyvodit žádná rozhodovací pravidla. Rozhodovací stromy se dají využít v případě, kdy má být provedena klasifikace nebo nějaká předpověď. Nejvhodnější je, je využívat v situaci, kdy lze hodnoty rozdělit do relativně malého počtu skupin. Rozhodně není vhodné je využívat v situaci, kdy je zapotřebí předpověď kvantitativních hodnot.

Strom tvoří jednotlivé uzly. Uzel na nejvyšší úrovni se nazývá kořenový. Veškeré vnitřní uzly představují testy jednotlivých atributů. I kořenový uzel je testem. Možný výsledek testu představuje větvení. Uzly, které jsou označovány jako listy, jsou listy, které se dále již nedělí. Takovéto uzly reprezentují jednotlivé třídy. [1, 22]

Cílem této metody je tedy nalézt takový stroj, který je konzistentní s trénovacími daty. Zpravidla se upřednostňují menší a jednodušší stromy. Obecné schéma algoritmu pro vytváření rozhodovacího stromu zobrazuje obrázek 5.

TDIDT algoritmus

1. zvol jeden atribut jako kořen dílčího stromu,
2. rozděl data v tomto uzlu na podmnožiny podle hodnot zvoleného atributu a přidej uzel pro každou podmnožinu,
3. existuje-li uzel, pro který nepatří všechna data do téže třídy, pro tento uzel opakuj postup od bodu 1, jinak skonči.

Obrázek 5: Obecný algoritmus pro tvorbu rozhodovacího stromu

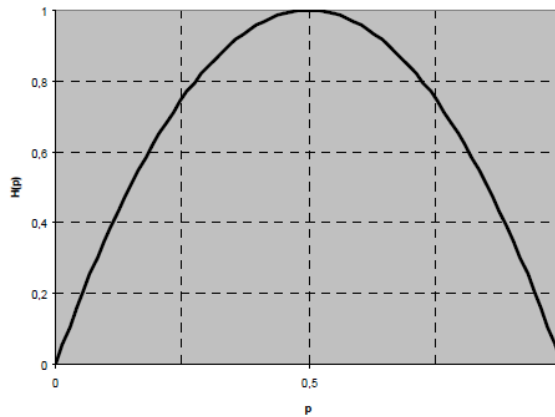
Zdroj:[1]

Tento algoritmus je použitelný pro kategoriální data, která nejsou zatížena šumem. Počet podmnožin vytvářených v kroku 2 se rovná počtu hodnot daného atributu. Růst stromů se podle bodu 3 zastaví v okamžiku, kdy všechny příklady v daném uzlu patří do téže třídy. Hlavní problém tohoto algoritmu je, jak vybrat vhodný atribut pro větvení stromu (bod1). Cílem je tedy zvolit atribut, který od sebe nejlépe odliší příklady různých tříd. Vodítkem pro výběr jsou charakteristiky atributu převzaté z teorie informace nebo pravděpodobnosti: entropie, informační zisk, poměrný informační zisk nebo Gini index. [1]

Entropie je termín, který se využívá v přírodních vědách jako například fyzika. Používá se pro vyjádření míry neuspořádanosti nějakého systému. V teorii informace je entropie definovaná jako funkce

$$H(p) = - \sum_{t=1}^T (p_t \log_2 p_t) \quad (5)$$

kde p_t představuje pravděpodobnost výskytu třídy t (v našem případě relativní četnost třídy t počítané na určité množině příkladů) a T je počet tříd. Grafické zobrazení průběhu entropie představuje obrázek 6. [1]



Obrázek 6: Průběh entropie

Zdroj:[1]

Pro vytváření rozhodovacích stromů a pravidel existuje množství algoritmů. Mezi nejznámější patří algoritmus CART, CLS, ID3, GID3, C4.5, AID, TREEDISC, QUEST a CHAID.

➤ ID3

Problematika algoritmu ID3 je popsána v následujícím obrázku 7:

1. Strom začíná jako samostatný uzel reprezentující tréninkové záznamy
2. Pokud jsou všechny záznamy stejné třídy, potom se tento uzel stane listem a je označen touto třídou
3. Jinak algoritmus použije metodu označovanou jako „informační přínos“ k určení atributu, který nejlépe rozdělí vzorky do jednotlivých tříd. Tento atribut se stane buď „testem“ nebo „rozhodnutím – třídou“. Tato verze algoritmu pracuje pouze s diskrétními hodnotami. Proto musí být spojité hodnoty diskretizovány.
4. Pro každou možnou hodnotu atributu je vytvořena samostatná větev a vzorky jsou přiřazeny jednotlivým větvím.
5. Algoritmus stejným způsobem vytváří rozhodovací strom v každé části. Jakmile se atribut použije jako testové kritérium, nesmí být znovu použit pro testování v této části stromu.
6. Opakování dělení může skončit, pokud je splněna některá z následujících podmínek:
 - všechny vzorky daného uzlu patří do stejné třídy,
 - nejsou žádné zbývající atributy, podle kterých by mohlo proběhnout další dělení. V tomto případě proběhne tzv. „většinová volba“. Ta zahrnuje převedení daného uzlu na list a označí se většinovou třídou ve vzorcích,

-nejsou žádné další vzorky patřící k dané větvi. V tomto případě je list vytvořen z většinové třídy ve vzorcích.

Obrázek 7: Algoritmus ID3

Zdroj:[22]

- **CART**-klasifikační a regresní strom. Úplný binární stromový algoritmus pro postupné binární štěpení datového souboru a tvorbu homogenních podmnožin.
- **QUEST**- statistický algoritmus pro selekci proměnných bez vychýlení. Sestavuje přesné binární stromy rychle a účinně
- **CHAID**-rychlý statistický víceúrovňový stromový algoritmus pro účinné zkoumání interakčních vztahů v datech. [1, 22]

2.3.Asociační pravidla

Asociační pravidlo (Association Rule – AR) představuje IF-THEN konstrukci. Tyto IF-THEN konstrukce lze nalézt ve všech programovacích jazycích, ale používají se i v běžné mluvě (nebude-li zima, vyrazíme ven). Z tohoto důvodu tedy pravidla s touto syntaxí radíme společně s rozhodovacími stromy k nejčastěji používaným prostředkům pro reprezentaci znalostí. A to ať už získaných od expertů, nebo vytvořených automatizovaně z dat. Pravidla IF-THEN zpravidla používáme pro klasifikaci. Při použití pravidel pro klasifikaci je závěr pravidel vyhrazen pro cílový atribut určující zařazení do třídy. [1,23]

Pojem asociační pravidlo bylo zpopularizováno v 90. letech 20. století R. Agrawalem. Agrawal představil asociační pravidla v souvislosti s analýzou nákupního košíku. Cílem celé analýzy bylo nalézt taková pravidla, která odpovídají implikaci: „Když zákazník koupí zboží X, tak koupí i zboží Y.“ Cílem analýzy se tedy zjišťuje, jaké druhy zboží si současně kupují zákazníci v supermarketech (např. alkohol a uzeniny). Jedná se tedy o hledání vzájemných vazeb, tak zvaných asociací, mezi různými položkami sortimentu prodejny. [23]

U pravidel vytvořených z dat nás obvykle zajímá, kolik příkladů splňuje *předpoklad* (**Ant-antecedent**) a kolik *závěr* (**Suc-succedent**) pravidla. Dále kolik příkladů splňuje předpoklad i závěr současně, kolik příkladů splňuje předpoklad a nesplňuje závěr atd. Tyto vztahy lze vyjádřit takto:

Ant => Suc,

Antecedent představuje levou stranu pravidla a sukcedent pravou stranu pravidla. Předpoklad a závěr představují kombinace kategorií, vypadá příslušná kontingenční tabulka. Zjednodušená kontingenční tabulka o rozměrech 3×3 se nazývá čtyřpolní tabulka. Taková to čtyřpolní tabulka je zobrazena níže v tabulce 1. Tato tabulka je doplněna ještě o řádkové a sloupcové sumy.

Tabulka 1: Čtyřpolní kontingenční tabulka

	závěr	¬závěr	Σ
předpoklad	<i>a</i>	<i>b</i>	$r = a + b$
¬předpoklad	<i>c</i>	<i>d</i>	$s = c + d$
Σ	$k = a + c$	$l = b + d$	$n = a + b + c + d$

Zdroj: upraveno podle [1]

Základní charakteristiky asociačních pravidel, jsou *podpora (support)* a *spolehlivost (confidence)*. Tyto dvě základní charakteristiky lze vypočítat z výše uvedené čtyřpolní kontingenční tabulky. Podpora je (absolutní, popř. relativní) počet objektů splňujících předpoklad i závěr. Podporu lze vyjádřit takto [1,23]:

$$a, \text{ resp. } P(\text{Ant} \wedge \text{Suc}) = \frac{a}{a+b+c+d}, \quad (6)$$

kde *a* představuje počet záznamů, které splňují předpoklad i závěr současně a *b* představuje počet záznamů, které splňují předpoklad a nesplňují závěr. Podporu lze také vyjádřit jako [23]:

$$\text{sup}(X \rightarrow Y) = \frac{\text{počet košíků obsahujících } X \text{ i } Y}{\text{počet všech košíků}} \quad (7)$$

Spolehlivost udává, jak moc se na dané pravidlo můžeme spolehnout a vypočítá se pomocí následující rovnice [1]:

$$P(\text{Suc} \mid \text{Ant}) = \frac{a}{a+b} \quad (8)$$

Spolehlivost (konfidence) lze dále vyjádřit ve tvaru:

$$\text{conf}(X \rightarrow Y) = \frac{\text{počet košíků obsahujících } X \text{ i } Y}{\text{počet košíků obsahujících } X} \quad (9)$$

Asociační pravidla se z dat získávají pomocí různých algoritmů. Nejčastěji se však jedná o algoritmus APRIORY. Vstupem algoritmu je vhodná reprezentace datové struktury a

parametry $minconf^6$ a $minsup^7$. Výstup představují všechny asociační pravidla $X \rightarrow Y$ taková, že $conf(X \rightarrow Y) \geq minconf$ a $sup(X \rightarrow Y) \geq minsup$. Podstatou algoritmu je, že nejprve vyhledá všechny frekventované podmnožiny položek k , které se v datech vyskytují dostatečně často (tj. s podporou $minsup$) a z těchto sestaví asociační pravidla splňující $conf(X \rightarrow Y) \geq minconf$ [1,23].

Algoritmus apriory:

1. do L_1 přiřad' všechny kategorie, které dosahují alespoň požadované četnosti
2. polož $k=2$
3. Dokud $L_{k-1} \neq \emptyset$

3.1 pomocí funkce apriori-gen vygeneruj na základě L_{k-1} množinu kandidátů C_k

3.2 do L_k zařad' ty kombinace z C_k , které dosáhly alespoň požadovanou četnost

3.3 zvětš' počítadlo k

kde L_1 je frekventovaná množina o velikosti 1, k označuje položky frekventované množiny, C_k označuje množinu kandidátů. [1]

Funkce apriori-gen (L_{k-1}):

1. Pro všechny dvojice kombinací $Comb_p, Comb_q$ z L_{k-1}
 - 1.1 Pokud $Comb_p$ a $Comb_q$ se shodují v $k-2$ kategoriích, přidej $Comb_p \wedge Comb_q$ do C_k
2. Pro každou kombinaci $Comb$ z C_k
 - 2.1 pokud některá z jejich podkombinací délky $k-1$ není obsažena v L_{k-1} odstraň $Comb$ z C_k .

Kde $Comb_p$ a $Comb_q$ jsou kombinace předpokladu a závěru. Jádrem algoritmu je hledání často se opakujících množin položek (frequent itemsets). Jedná se o kombinace (konjunkce) kategorií, které dosahují předem zadané četnosti (podpory $minsup$) v datech. Při hledání kombinací délky k , které mají vysokou četnost, se využívá toho, že již známe kombinace délky $k-1$. Při vytváření kombinace délky k , spojujeme kombinace délky $k-1$. [23]

⁶ $minconf$ – označení pro minimální spolehlivost pravidla

⁷ $minsup$ – označení pro minimální podporu pravidla

3. Výběr datového souboru

Při výběru vhodných dat pro vytvoření výsledného modelu je nezbytně nutné dobré porozumění trhu a dobře si stanovit a promyslet vlastní cíle. Data při modelování slouží jako informační základna nebo jako rámec. Kvalita a relevantnost modelu velmi úzce souvisí s kvalitou zdrojových dat. Klíčem k vytvoření kvalitního a vypovídajícího modelu je odpovídající výběr dat a ověření cíleného modelu. Z hlediska toho, pro jaké účely mají být data využívána, rozlišujeme například data pro získávání zákazníků, data pro zákaznické modely, data pro modely rizika a jiné. [17]

3.1. Popis vstupních dat

Vstupní data využitá pro analýzu a modelování pocházejí ze dvou zdrojů. První soubor vstupních dat je převzat ze Sociologického ústavu AV ČR, v.v.i.⁸. Jedná se tedy o data z oblasti sociální politiky. Datový soubor byl získán prostřednictvím standardního dotazníku, který je přiložen v příloze A. Tento dotazník byl upraven pro potřeby analýz a modelování v diplomové práci. Datový soubor obsahuje 1294 záznamů a 80 proměnných. Sběr dat probíhal v období od začátku roku 1997 a končil současně s koncem roku 1997. Zájmové území bylo vymezeno na Českou republiku. Datový soubor obsahuje data od mladých svobodných lidí ve věku 18-30 let. Otázky v dotazníku jsou směřovány tak, aby bylo zjištěno co nejvíce informací o rodinném a reprodukčním chování mladých dosud svobodných lidí, o postoji k manželství a rodičovství. Dále byla prostřednictvím dotazníku zjišťována problematika vnějších podmínek, za kterých se mladí lidé rozhodují o uzavření manželství a narození dětí a plány mladé generace týkající se demografického chování. Celý výzkum byl realizován pod vedením RNDr. Fialové. [28]

Druhý datový soubor byl získán prostřednictvím vlastního dotazníkového šetření. Pro následné porovnávání výsledků, byl převzat identický dotazník (příloha A) z AV ČR, v.v.i, kterým byla získána data z prvního datového souboru. Tento dotazník byl rozšířen mezi vysokoškolsky vzdělané nebo stále na vysoké škole studující mladé lidi, také z věkové kategorie 18-30 let. Zájmové území bylo opět vytyčeno na celou Českou republiku. Data byla sbírána od začátku října 2011 do začátku roku 2012. Následně byla data z dotazníku zpracována a převedena do programového prostředí Microsoft Excel. Tím byla získána druhá

⁸ AV ČR v.v. i. -Akademie věd České republiky, veřejná výzkumná instituce

datová matice. Pro využití dat v programu Clementine bylo nutné data převést do formátu CSV (oddělené čárkou). Soubor byl tedy uložen s koncovkou *.csv. Soubor v takovém to formátu je již možné importovat do prostředí programu Clementine.

3.2. Analýza vstupních dat

Data získaná s datového archivu AV ČR bylo nutné nejprve upravit v programovém prostředí MS Excel. Bylo nutné upravit názvy proměnných pro potřeby následné analýzy a modelování. V prostředí MS Excel byl vytvořen i datový slovník, který je zobrazen v tabulce 3. Součástí datového slovníku je název atributu, typ proměnné (kategorická, spojitá), typ proměnné v programu Clementine (Set, Order Set, Range, Flag) a rozsah dat. Před zahájením dotazníkového šetření bylo nutné dotazník upravit. Byla vyřazena řada otázek, které nebyly nikterak důležité pro další práci. Z prvotního velkého množství proměnných bylo ponecháno pouze 80 proměnných, které dále představovaly vstupní data našeho řešeného problému. Popis jednotlivých atributů je následující:

- *Politická orientace* - první atribut se zabývá politickou orientací respondentů. Jedná se o zjištění, zda se respondent spíše přiklání k „pravici“ či „levici“.
- *Vývoj v ČR* – podstatou je zjištění pohledu respondent na vývoj ČR. Zda se vyvíjí všeobecně správným nebo nesprávným směrem.
- *Spokojenost se životem* - vzhledem ke všem okolnostem, spokojenost či nespokojenost se životem respondenta.
- *Rodina dětství (dospívání)* – typ rodiny, ve které respondent žil v dětství (do 12 let) resp. v dospívání (12-18) let. V úplně rodině s oba rodiči, v neúplné rodině pouze s matkou/otce, aj.
- *Životní úroveň dětství (dospívání)* – zhodnocení celkové životní úrovně, ve které respondent v dětství (do 12 let) resp. v dospívání (12-18 let) žil.
- *Sourozenci dětství (dospívání)* – počet sourozenců, se kterými respondent v období dětství, resp. dospívání žil.
- *Obtížné životní situace* – jakým způsobem obvykle respondent řeší nově vzniklé životní situace.

- *Povaha změny* – zjištění povahy a náтуры respondenta, z ohledu na to, jak přijímá změny v životě.
- *Atributy 12-18* – otázka, která se skládá z těchto atributů je hodnocena pěti bodovou stupnicí, kdy 1 vyjadřuje nejmenší důležitost a 5 největší důležitost pro respondenta. Soubor těchto atributů se týká osobní důležitosti pro respondenta. Jedná se o důležitost vzdělání, kvalitní práce, čas na koníčky a rodinu, čas na přátele, peníze aj.
- *Atributy 19-27* – otázka složená z těchto atributů je také bodovací. Tedy 1 vyjadřuje nejmenší důležitost a 5 největší důležitost. Soubor těchto atributů se zabývá důležitostí ohledně vlastností a charakteru o životním partnerovi. Respondent stanovuje pomocí bodovací stupnice, jak jsou pro něho důležité vlastnosti jeho životního partnera. Jedná se např. o vzdělání, vzhled, finanční zabezpečení, zodpovědnost, poctivost, sportovní založení, vztah k dětem, inteligence aj.
- *Atributy 28-34* – tento soubor atributů pojednává o důležitosti respondenta v situaci, kdy by se rozhodl uzavřít manželství. Respondent hodnotí jednotlivé atributy kategoriemi určitě ano (UA), spíše ano (SA), spíše ne (SN), určitě ne (UN). Respondent tedy hodnotí, jak důležitou roli by při rozhodování o uzavření manželství hrálo např. nesouhlas rodičů se sňatkem, možnost seberealizace, nutnost narození dětí v manželství, touha bezpečného zázemí, finanční zajištění aj.
- *Životní styl* – jaký styl života respondent spíše preferuje. Zda žít sám bez stálého partnera, žít se stálým partnerem bez sňatku, žít s partnerem nějaký čas a později uzavřít sňatek či uzavřít sňatek s partnerem bez předchozího společného soužití.
- *Muž (žena) věk 1. manželství* – věk, v němž by podle respondenta měl muž resp. žena uzavírat poprvé manželství.
- *Muž (žena) věk 1. resp. poslední dítě* – věk, v němž by si muž resp. žena měli pořizovat první resp. poslední dítě.
- *Ideální počet dětí* – ideální počet dětí v rodině podle respondenta
- *Atributy 43-49* – tyto atributy se týkají problematiky uzavírání sňatku. Jedná se opět o bodovací otázku, ve které 1 představuje nejmenší důležitost a 5 největší důležitost. Jednotlivé atributy představují situace týkající se uzavírání sňatku. Respondent hodnotí, jak jsou pro něho jednotlivé situace důležité v případě, že by uzavíral sňatek.

Jedná se například o dovršení plnoletosti, dokončení vzdělání, finanční samostatnost, samostatné bydlení, dlouhodobá známost s partnerem aj.

- *Plánování dětí* – budoucí plány ohledně dětí. Zda respondent děti plánuje, zatím neplánuje nebo děti nikdy mít vůbec nechce.
- *Počet plánovaných dětí*
- *Atributy 52-57* – tento soubor atributů pojednává problematice narození dítěte. Jedná se opět o bodovací otázku. Respondent hodnotí jednotlivé atributy podle toho, jakou mírou jsou pro něho důležité při rozhodnutí o pořízení dítěte. Jedná se o dovršení plnoletosti, dokončení vzdělání, dostatečný finanční příjem, samostatné bydlení, dlouhodobá známost s partnerem aj.
- *Atributy 58-68* – otázka, která se skládá z těchto atributů je hodnotící. Respondent využívá kategorie: určitě souhlasím (US), spíše souhlasím (SS), spíše nesouhlasím (SN), určitě nesouhlasím (UN). Hodnotí se jednotlivé výroky jako např. : hlavní výhodou manželství je finanční jistoty; hlavním uzavíráním sňatků je mít děti; lidé, kteří by chtěli mít děti, mají uzavřít sňatek; sledovat, jak děti rostou, je největší radost v životě; manželství vždy omezuje svobodu ženy i muže; nezaměstnanost se v naší zemi bude stále více zvyšovat; každý pracovitý člověk si v dnešní době může v naší zemi hodně vydělat aj.
- *Životní minimum pro 3-člennou rodinu* – peněžní částka v české měně (CZK). Respondent stanovuje, kolik si myslí, že je v dnešní době životní minimum pro tříčlennou rodinu.
- *Rozvody* – názor respondenta na problematiku rozvodů. Respondent volí mezi výroky: rozvod by měl být jednoduchou záležitostí; pokud jsou v rodině děti, rozvod by měl být nesnadný, aby se udrželo rodinné prostředí; rozvod by neměl být snadnou záležitostí pro žádné manželství; rozvod by měl být zakázán.
- *Pohlaví respondentů*
- *Sociální postavení* – sociální postavení představuje aktuální situaci respondenta: student, nezaměstnaný, zaměstnaný, družstevník, živnostník, podnikatel, ostatní.
- *Vzdělání* – nejvyšší ukončené školní vzdělání. Základní bez vyučení, vyučen bez maturity, maturita, vysoká škola.

- *Student* – zda je v současné době respondent studentem vysoké školy či již ne.
- *Čistý měsíční příjem* – odhad čistého měsíčního příjmu respondenta. Hodnoty jsou uvedeny v české měně (CZK)
- *Dluhy* – odhad výše finančních dluhů respondenta. Hodnoty jsou uvedeny v české měně (CZK).
- *Velikost obce* – respondent volí mezi kategoriemi velikosti obce, ve které má trvalé bydliště. Jednotlivé kategorie jsou uvedeny v dotazníku, který je přílohou A, v otázce číslo 26.
- *Území* – území ČR bylo rozděleno do 8 území. Respondent volí, ve kterém území má trvalé bydliště. Praha, Střední Čechy, Jižní Čechy, Západní Čechy, Severní Čechy, Východní Čechy, Jižní Morava, Severní Morava.
- *Věková kategorie* – rozdělení věku 18-30 let do 4 věkových kategorií. Jednotlivé kategorie jsou uvedeny v dotazníku (příloha A) otázka číslo 28.
- *Čistý měsíční příjem* – stanovení 5-ti kategorií čistého měsíčního příjmu v české měně (CZK), ze kterých respondent vybírá. Rozdělení do jednotlivých kategorií je uvedeno v dotazníku (příloha A) poslední otázka číslo 29.

Z důvodu, že dotazníkové šetření je spíše kvantitativním výzkumem, meritorními otázkami jsou většinou otázky „uzavřené“. **Meritorní** otázky představují otázky, jež jsou zaměřeny na problémy, které jsou předmětem výzkumu. U uzavřených otázek respondent vybírá pouze z předepsaných variant odpovědí. Tyto otázky jsou velmi výhodné z hlediska zpracování a dále i vyhodnocování odpovědí, kdy lze využít různých statistických metod. Problém, který s nimi ale souvisí, je ten, že se respondentům vlastně vnucují určité varianty odpovědí. Tím dochází ke zkreslování údajů. Z praxe a zkušenosti je však známé, že respondenti mají raději „uzavřené“ otázky než otázky „otevřené“. V dotazníku pro tuto práci se jedná o větší část otázek, například o otázky 1,2,3 aj. Nemusejí tak nijak zvláště přemýšlet nad odpovědí. V malé míře jsou v dotazníku uvedeny i otázky „polozavřeného“ typu. Respondenti si mohou vybrat buď některou z předepsaných variant odpovědi či formulovat vlastní odpověď. Ze zkušeností, ale většinou vybírají některou z předepsaných variant. Uzavřené otázky, které umožňují výběr z většího množství variant, se nazývají selektivní. Jednotlivé varianty je vhodné volit tak, aby odpovědi diferencovaly respondenty z jednoho hlediska. [7]

Součástí dotazníku nejsou pouze meritorní otázky, ale také **analytické otázky**. Ty slouží především při zpracování primárních údajů, aby mohli být respondenti tříděni do skupin a aby mohly být analyzovány zvláštnosti těchto skupin. Součástí jich jsou identifikační otázky. Identifikační otázky slouží k popisu nejdůležitějších vlastností vybraných osob, firem či jiných vybraných jednotek. Jedná se například o pohlaví, vzdělání, stáří, předmět podnikání, počet zaměstnanců apod. V dotazníku této práce se jedná například o otázky číslo 21, 22 či 23. [7]

Velmi důležitým typem otázek v dotazníku jsou “**kontrolní**” otázky. Díky těmto otázkám se ověřuje správnost odpovědí na otázky, u nichž se předpokládá záměrné zkreslování odpovědí. Kontrolní otázky by měli být sestavovány a umístěny v dotazníku tak, aby bylo jen velmi obtížné neodhalit jejich funkci. [7]

Tabulka 2: Datový slovník

Atribut	Typ proměnné	Typ v Clementine	Rozsah dat
<i>politicka_orientace</i>	kategoriální	Set	[1-5]
<i>vyvoj_v_CR</i>	kategoriální	Set	[1-3]
<i>spokojenost_se_zivotem</i>	kategoriální	Set	[1-4]
<i>rodina_detstvi</i>	kategoriální	Set	[1-7]
<i>rodina_dospivani</i>	kategoriální	Set	[1-7]
<i>zivotni_uroven_detstvi</i>	kategoriální	Set	[1-4]
<i>zivotni_uroven_dospivani</i>	kategoriální	Set	[1-4]
<i>sourozenci_detstvi</i>	spojitá	Range	
<i>sourozeni_dospivani</i>	spojitá	Range	
<i>obtizne_zivotni_situace</i>	kategoriální	Set	[1-4]
<i>povaha_zmeny</i>	kategoriální	Set	[1-4]
<i>dulezitest_vzdelani</i>	kategoriální	Order Set	[1-5]
<i>dulezitest_cas_na_konicky</i>	kategoriální	Order Set	[1-5]
<i>dulezitest_prace</i>	kategoriální	Order Set	[1-5]
<i>dulezitest_rodinny_zivot</i>	kategoriální	Order Set	[1-5]
<i>dulezitest_uspechy_a_uznani</i>	kategoriální	Order Set	[1-5]
<i>dulezitest_cas_na_pratele</i>	kategoriální	Order Set	[1-5]
<i>dulezitest_hodne_penez</i>	kategoriální	Order Set	[1-5]
<i>partner_VŠ_vzdelani</i>	kategoriální	Order Set	[1-5]
<i>partner_finance</i>	kategoriální	Order Set	[1-5]
<i>partner_vzhled</i>	kategoriální	Order Set	[1-5]
<i>partner_sportovni_zalozeni</i>	kategoriální	Order Set	[1-5]
<i>partner_zodpovednost_a_poctivost</i>	kategoriální	Order Set	[1-5]
<i>partner_vztah_k_rodine_a_detem</i>	kategoriální	Order Set	[1-5]
<i>partner_inteligence</i>	kategoriální	Order Set	[1-5]
<i>partner_tolerantnost</i>	kategoriální	Order Set	[1-5]

Atribut	Typ proměnné	Typ v Clementine	Rozsah dat
<i>partner_prizpusobivost</i>	kategoriální	Order Set	[1-5]
<i>Manzelstvi_nesouhlas_rodicu</i>	kategoriální	Order Set	[1-4]
<i>Manzelstvi_deti</i>	kategoriální	Order Set	[1-4]
<i>Manzelstvi_moznost_seberealizace</i>	kategoriální	Order Set	[1-4]
<i>Manzelstvi_bezpecne_zazemi</i>	kategoriální	Order Set	[1-4]
<i>Manzelstvi_pravidelny_sex</i>	kategoriální	Order Set	[1-4]
<i>Manzelstvi_financi_zajisteni</i>	kategoriální	Order Set	[1-4]
<i>Manzelstvi_jasne_majetkove_vztahy</i>	kategoriální	Order Set	[1-4]
<i>zivotni_styl</i>	kategoriální	Set	[1-4]
<i>muz_vek-1.manzelstvi</i>	spojitá	Range	
<i>zena_vek-1.manzelstvi</i>	spojitá	Range	
<i>muz_vek-1.dite</i>	spojitá	Range	
<i>zena_vek-1.dite</i>	spojitá	Range	
<i>muz_vek-1.posledni_dite</i>	spojitá	Range	
<i>zena_vek-1.posledni_dite</i>	spojitá	Range	
<i>idealni_pocet_deti</i>	spojitá	Range	
<i>snatek_plnoletost</i>	kategoriální	Order Set	[1-5]
<i>snatek_ukoncene_vzdelani</i>	kategoriální	Order Set	[1-5]
<i>snatek_ekonomicka_nezavislost</i>	kategoriální	Order Set	[1-5]
<i>snatek_finance</i>	kategoriální	Order Set	[1-5]
<i>snatek_uspory</i>	kategoriální	Order Set	[1-5]
<i>snatek_samostatne_bydleni</i>	kategoriální	Order Set	[1-5]
<i>snatek_dlouha_znamost</i>	kategoriální	Order Set	[1-5]
<i>planovani_deti</i>	kategoriální	Set	[1-3]
<i>počet_planovanych_deti</i>	spojitá	Range	
<i>dite_plnoletost</i>	kategoriální	Order Set	[1-5]
<i>dite_ukoncene_vzdelani</i>	kategoriální	Order Set	[1-5]
<i>dite_finance</i>	kategoriální	Order Set	[1-5]
<i>dite_uspory</i>	kategoriální	Order Set	[1-5]
<i>dite_samostatne_bydleni</i>	kategoriální	Order Set	[1-5]
<i>dite_dlouha_znamost</i>	kategoriální	Order Set	[1-5]
<i>financni_vyhoda_manzelstvi</i>	kategoriální	Order Set	[1-4]
<i>deti_jako_hlav_duvod_manzelstvi</i>	kategoriální	Order Set	[1-4]
<i>kdo_chce_deti_mel_by_se_ozenit/vdat</i>	kategoriální	Order Set	[1-4]
<i>deti_nejvetsi_zivotni_radost</i>	kategoriální	Order Set	[1-4]
<i>deti_omezeni_svobody_rodicu</i>	kategoriální	Order Set	[1-4]
<i>bezdetny_zivot_je_prazdny</i>	kategoriální	Order Set	[1-4]
<i>manzelstvi_omezuje_svobodu</i>	kategoriální	Order Set	[1-4]
<i>zlepseni_zivotni_urovne</i>	kategoriální	Order Set	[1-4]
<i>zlepseni_dostupnosti_bydleni_pro_mlade</i>	kategoriální	Order Set	[1-4]
<i>pracovity_clovek_muze_vydelat_penize</i>	kategoriální	Order Set	[1-4]
<i>nezamestnanost_se_bude_zvysovat</i>	kategoriální	Order Set	[1-4]
<i>život_min_pro_3_clenou_rodinu</i>	spojitá	Range	
<i>rozvody</i>	kategoriální	Set	[1-4]

Atribut	Typ proměnné	Typ v Clementine	Rozsah dat
<i>pohlavi_respondentu</i>	kategoriální	Flag	[1;2]
<i>socialni_postaveni</i>	kategoriální	Set	[1-7]
<i>vzdelani</i>	kategoriální	Set	[1-4]
<i>student</i>	kategoriální	Flag	[1;2]
<i>cisty_mesicni_prijem</i>	spojitá	Range	
<i>uspory</i>	spojitá	Range	
<i>velikost_obce</i>	kategoriální	Set	[1-5]
<i>uzemi</i>	kategoriální	Set	[1-8]
<i>vekova_kategorie</i>	kategoriální	Set	[1-4]
<i>cisty_mesicni_prijem_kategorie</i>	kategoriální	Set	[1-5]

Zdroj: vlastní zpracování

4. Návrh modelů vybrané oblasti sociální politiky

Následující čtvrtá kapitola této práce je věnována nejprve přípravě dat a poté vlastnímu modelování. Nejprve bylo nutné připravit data pro vybrané metody modelování. Postup úpravy dat je popsán níže v kapitole příprava dat a modelování. Po samotné přípravě dat byly použity některé z vybraných metod modelování. Postup tvorby jednotlivých modelů, včetně grafického zobrazení výsledků, komparace výsledků a zhodnocení výsledků, je uvedeno níže v této kapitole.

4.1. Příprava dat a modelování

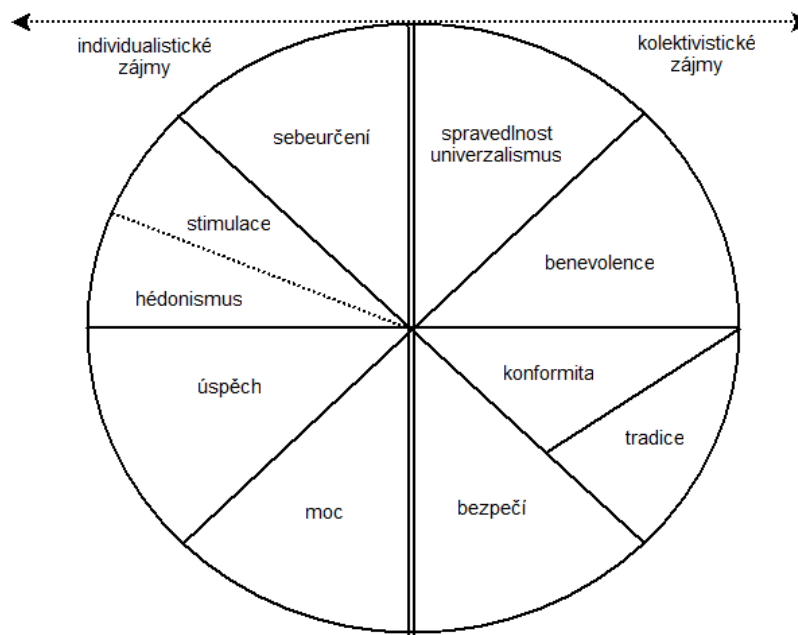
Před zahájením samotného modelování a tedy použití vybraných metod modelování na data bylo nutné vyřešit problém chybějících hodnot v datovém souboru. Problematika chybějících hodnot se vyskytuje téměř v každé sadě dat. Je důležité skutečnost, že hodnota chybí neignorovat, ale vždy je nutné tuto skutečnost zachytit. Skutečnost, že hodnota chybí, totiž může mít prediktivní vlastnosti. Existují tři základní způsoby, jak řešit problém chybějících hodnot: substituce jedné hodnoty, substituce střední hodnotou třídy a regresní substituce. Nejjednodušší metodou nahrazování chybějících hodnot je substituce jedné hodnoty. Zde obvykle pracujeme se střední hodnotou, mediánem nebo módem. V případě, že rozdělení dat není špičaté (souměrné), využíváme střední hodnotu. Střední hodnot je založena na statistickém výpočtu nejmenší chyby čtverců. Je-li rozdělení velmi špičaté, může lépe posloužit medián. Při substituci střední hodnotou třídy se využívají střední hodnoty podskupin jiných proměnných nebo kombinací proměnných. Tato metoda zachovává lépe původní rozdělení hodnot. Regresní substituce střední hodnotou třídy využívá regresní substituce střední hodnoty skupin jiných proměnných. Velkou výhodou regrese je schopnost pracovat se spojitými proměnnými stejně jako hledat ve více proměnných přesnější míru. Výsledné hodnocení regrese slouží k dopočtení náhradních hodnot. [17]

Při přípravě obou datových souborů pro modelování byla zvolena možnost substituce jedné hodnoty. Z důvodu, že je nerovnoměrné rozdělení dat, byl pro nahrazení chybějících hodnot využit medián. Tím tedy byly nahrazeny chybějící hodnoty u obou datových souborů. Záznamy, ve kterých byl velký počet chybějících hodnot, byly zcela z datového souboru vyloučeny. Problematiku odlehlých hodnot nebo chyb v datech nebylo nutné v ani jednom souboru řešit. Řešení nahrazení chybějících hodnot, je uvedeno v příloze B. Součástí této přílohy, je také základní statistika souboru.

4.2. Model komparace preferencí politických stran voliče v roce 1997 a 2011

Data, která budou v tomto modelu zpracovávána, se týkají vysokoškolsky vzdělaných nebo stále na vysoké škole studujících lidí ve věku 18-30 let. Zájmové území bylo vytyčeno pouze na oblast východních Čech. Pro analýzu i následné modelování, budou použity oba datové soubory, které jsou popsány výše.

U politické orientace jsou rozlišovány dvě relativně nezávislé dimenze. První z těchto dvou dimenzí je levo/pravou dimenzí. Následná analýza a modelování na datech bude vycházet právě z tohoto typu dimenze. Podstatou problému bude tedy analyzovat, jak se změnily preference politických stran (levice, pravice) voličů od roku 1997 do roku 2011. Tato první dimenze vyjadřuje podle názoru některých autorů postoj k tržnímu rozdělování a státnímu přerozdělování, který je odvozován od hodnoty, jíž je rovnost. Podrobnější definování termínů hodnota a postoj jsou uvedeny u druhého modelu. Podle Dahrendofra je v pozadí této dimenze základní moderní sociální konflikt, jímž je konflikt mezi požadavky inovace (hospodářského růstu) a spravedlnosti (sociální rovnosti). Druhou dimenzí politické orientace je dimenze liberalismu/autoritářství. Jde o dimenzi, jež je zčásti odvozena od známé F-škály. Tento typ dimenze nebyl v práci dále uvažován. Existuje tak zvaný Schwartzův systém hodnot. Tento systém vytvořili autoři Schwartz a Bilsky. Kruhové uspořádání terminálních hodnot podle Schwartze je zobrazeno na obrázku 8. Jedná se o teorii struktury a obsahu hodnot. Autoři vyšli z předpokladu, že všichni lidé jsou vystaveni působení tří základních typů požadavků: musí uspokojovat biologické potřeby, zúčastňovat se koordinovaných sociálních interakcí a dostát institucionalizovaným požadavkům vztahujícím se k možnosti skupinového přežití. Lze konstatovat, že mezi tímto Schwartzovým systémem hodnot a výše uvedeným dvojdimenzionálním pojetím politické orientace existuje vysoká míra shody. Levo/pravé dimenzi politické orientace odpovídá v Schwartzově modelu protiklad kvadrantů spravedlnost/benevolence – úspěch/moc. Dimenzi liberalismu/autoritářství odpovídá protiklad zbylých dvou kvadrantů hodnot. Klíčové protikladné hodnoty zde představují zejména sebeurčení (svoboda, nezávislost) – tradice/konformita. [16]



Obrázek 8: Kruhové uspořádání terminálních hodnot podle Schwartz

Zdroj:[16]

Analýza

Podstatou řešeného problému je analyzovat, jak se změnilы preference politických stran voličů od roku 1997 do roku 2011. Důvod, proč je tato problematika řešena, spočívá v tom, že objem vysokoškoláků, kteří se účastní voleb, je v porovnání s ostatními sociálními skupinami poměrně malý. Malé zastoupení vysokoškoláků proto nemusí mít tak vypovídající schopnosti o této skupině. Proto bylo realizováno dotazníkové šetření na tuto problematiku pouze mezi vysokoškoláky. V dotazníkovém šetření, byli lidé tázáni, zda se spíše podle svých názorů přiklání k levicovým či pravicovým stranám nebo zda si myslí, že patří spíše mezi nestranné, tedy střed. Byly porovnávány zjištěné výsledky v roce 1997 s výsledky zjištěnými na přelomu roku 2011 a 2012 (realizované dotazníkové šetření). V roce 1997 do levicových stran patřilo ČSSD, KSČM, DŽJ (Důchodci za životní jistoty). Pravicové strany byly ODS, KDU-ČSL, US (Unie svobody) a DEU (Demokratická unie). Během čtrnácti resp. patnácti let došlo pochopitelně ke změnám stran ať už levice či pravice, ale dominantní strany jako ČSSD, KSČM, ODS či KDU-ČSL stále zůstávají neměnné. Získané výsledky preferencí voličů od roku 1997 do roku 2011 jsou uvedeny v tabulce 3. Data jsou ve formátu procent (%).

Tabulka 3: Vývoj preference politických stran voliče od roku 1997 do roku 2011

	1997	2011
LEVICE	35,93	15,33
STŘED	17,96	21,33
PRAVICE	46,11	63,33

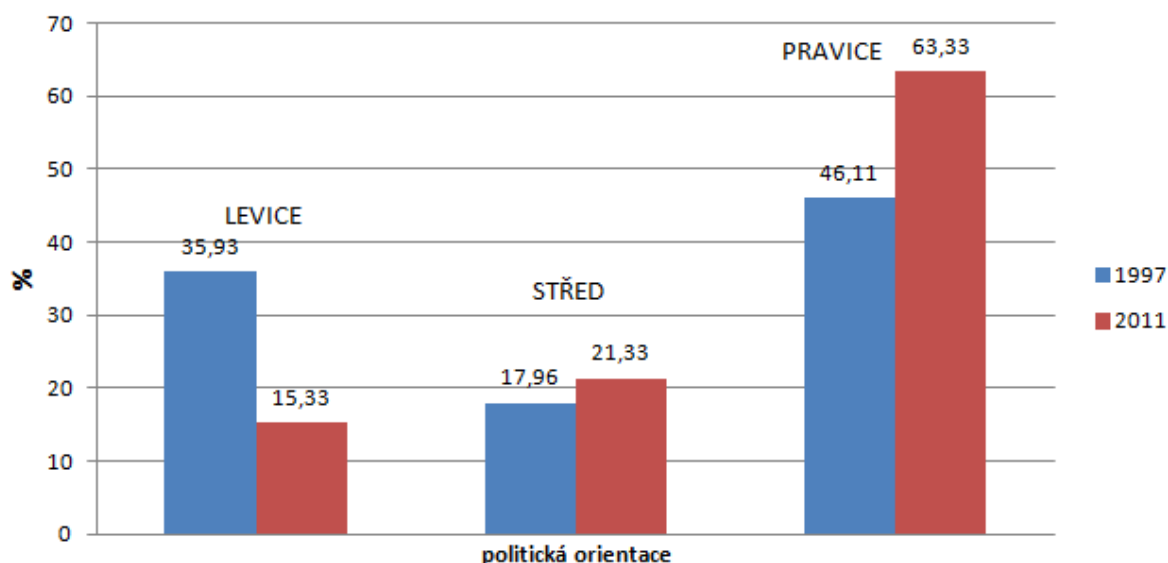
Zdroj: vlastní zpracování

Pravicové strany jsou většinou charakterizovány jako strany, které podporují zachování tradičních a kulturních hodnot a zvyklostí či zachování určité formy sociální hierarchie. Naopak levicové strany jsou strany podporující změnu tradičního sociálního pořádku a to s cílem vyrovnanějšího rozdělování bohatství a privilegií. [19]

Na základě získaných výsledků, které jsou uvedeny v tabulce 3 a graficky jsou znázorněny na grafu 1 lze konstatovat, že od roku 1997 do roku 2011 došlo k poměrně značné změně preferencí voličů. V roce 1997 je rozdíl mezi preferencemi levice a pravice poměrně malý a to zhruba 10%. Lze tedy říci, že poměr levicově a pravicově smýšlejících vysokoškoláků je poměrně vyrovnaný. To však nelze říci o výsledcích z roku 2011. Z tabulky 3 je patrné, že rozdíl mezi pravicí a levicí je značný. Nadpoloviční většina vysokoškoláků se přiklání spíše k pravicovým stranám a jen zhruba jedna šestina dotázaných volí levicové strany. Dalším důležitým faktem je, že stále zůstává poměrně vysoký objem vysokoškoláků, kteří se nepřiklání většinově ani k jedné straně a volí spíše střed. Z celkového objemu je to v obou letech zhruba jedna pětina vysokoškoláků.

Pokud porovnáme výsledky obou let, lze říci, že od roku 1997 do roku 2011 došlo ke značnému poklesu vysokoškoláků, kteří volí levicové strany a to zhruba o 20%. Naopak došlo k nárůstu, zhruba o 17%, vysokoškoláků, kteří spíše preferují pravicové strany. Objem nestranných stále zůstává na stejné úrovni, zhruba 20% z dotázaných.

Preference politické orientace pro rok 1997 a 2011



Graf 1: Preference politické orientace pro rok 1997 a 2011

Zdroj: vlastní zpracování

Z výsledků předčasných parlamentních voleb v roce 1998, které jsou uvedeny v tabulkách 4 a 5, je patrné, že rozdíl mezi pravicí a levicí je zanedbatelný. Lze říci, že preference voličů obou stran jsou vyrovnané. Nutno dodat, že tyto výsledky jsou obecné, pocházejí tedy od všech sociálních skupin, různých věkových kategorií a území celé ČR. Pokud tyto obecné výsledky porovnáme s výsledky vysokoškoláků 18-30 let, lze konstatovat, že jsou poměrně podobné. V letech 1997 resp. 1998 byla preference všech voličů, tedy ze všech sociálních skupin vyrovnaná. Voliči se tedy nepřikláněli spíše ani k levici ani k pravici.

Tabulka 4: Výsledky voleb pravice v roce 1998

PRAVICE	Výsledky voleb
ODS	27,7
KDU-ČSL	9
US	8,6
DEU	1,5
	46,8

Zdroj:[26]

Tabulka 5: Výsledky voleb levice v roce 1998

LEVICE	Výsledky voleb
ČSSD	32,3
KSČM	11
DŽJ	3,1
	46,4

Zdroj:[26]

Hodnoty v tabulkách jsou uvedeny v procentech (%). Malé strany, jejichž výsledky byly zanedbatelné (nižší než 1%), nebyli bráni v úvahu.

V tabulce 6 jsou uvedeny obecné výsledky voleb z roku 2010, z kterých je patrné, že větší objem hlasů získaly pravicově zaměřené strany. Pokud porovnáme výsledky parlamentních voleb z roku 2010 s výsledky od vysokoškoláků z roku 2011, rozdíly v preferencích stran jsou patrné. Zatímco obecné výsledky, tedy od všech sociálních skupin, ve všech věkových kategoriích a z celého území ČR vypovídají o vyrovnaných preferencích voličů, výsledky od vysokoškoláků nám ukazují na většinovou preferenci pravicově zaměřených stran. Na základě toho tedy lze konstatovat, že skupina vysokoškolsky vzdělaných lidí spíše preferuje pravicově zaměřené strany a většina ostatních sociálních skupin volí tedy spíše levicové strany. Na základě výzkumu sociologického ústavu AV ČR, lze konstatovat že: „Lze učinit závěr, že obliba levice klesá s růstem vzdělání. Pokles však není rovnoměrný, obzvláště velký rozdíl je ve volebním chování vyučených voličů a voličů s maturitou“. [26, str. 6-7]

Tabulka 6: Výsledky voleb v roce 2010

	Výsledky voleb
PRAVICE	48,62
LEVICE	39,30

Zdroj:[35]

Modelování

V následující části budou použity jednotlivé modelovací metody, kterou jsou popsány výše. Výstupem modelování se předpokládá proměnná, která vyjadřuje typického voliče levice, resp. pravice v roce 1997 a v roce 2011. Vzhledem k tomu, že výše uvedená analýza vývoje preferencí politických stran voliče se týká pouze vysokoškolsky vzdělaných lidí nebo

stále na vysoké škole studujících lidí a to pouze z Východních Čech, tak i modelování bude zaměřeno na tuto cílovou skupinu. Typickým voličem se tedy myslí volič vysokoškolák z Východních Čech. Obecný model pro řešení daného problému je uveden na obrázku 9. Jako vstupy modelování jsou zvoleny proměnné názor na vývoj v ČR, zde v dotazníku respondenti volili mezi možnostmi, ČR se vyvíjí správným směrem, vyvíjí se špatným směrem nebo možnost vývoj nesměruje nikam. Dalším zvoleným vstupem byla životní úroveň respondenta v dětství a dospívání. Respondent volil z možností životní úroveň velmi dobrá, spíše dobrá, spíše špatná a velmi špatná. Dalším vstupem do modelu bylo zvoleno pohlaví a sociální postavení. Sociálním postavením se myslí, zda respondent je studentem či zaměstnancem. Dále věková kategorie a čistý měsíční příjem. Z důvodu, že cílový segment je mladá generace ve věku 18 až 30 let, věk byl tedy rozdělen do kategorií také v tomto rozsahu. Při výběru vhodných vstupních proměnných do modelu, bylo vycházeno ze zdroje [9], kde jsou popsány silné resp. slabé faktory ovlivňující preference politických stran voličů.



Obrázek 9: Obecný model - typický volič levice/pravice v roce 1997 a 2011

Zdroj: vlastní zpracování

Jak je patrné z výše uvedeného obecného modelu, pro modelování budou použity dvě modelovací metody, jedná se o asociační pravidla a shlukovou analýzu. Problematika asociačních pravidel byla popsána výše v kapitole Vybrané metody modelování. V rámci shlukové analýzy bylo možné použít algoritmus K-means, Kohonenovy mapy a algoritmus TwoStep. Pro řešení daného byl vybrán algoritmus TwoStep. Základní charakteristiky těchto tří algoritmů jsou uvedeny níže v tabulce 7. Algoritmus TwoStep byl vybrán převážně z toho důvodu, že podporuje práci s kategorizovanými proměnnými, což algoritmus K-means nepodporuje.

Tabulka 7: Srovnání shlukovacích metod

	Silné stránky	Slabé stránky
Kohonenova mapa	<ul style="list-style-type: none"> • <i>Pro spojité i kategorizované proměnné</i> • <i>Nahradí chybějící hodnoty</i> • <i>jednoduchá</i> 	<ul style="list-style-type: none"> • <i>Výsledek závisí na subjektivní představě (navrhnout rozměry mapy)</i>
K-means	<ul style="list-style-type: none"> • <i>Jednoduchá a efektivní</i> • <i>Nejrychlejší způsob shlukování pro velké datové soubory</i> • <i>Chybějící hodnoty jsou nahrazeny hodnotou 0,5</i> 	<ul style="list-style-type: none"> • <i>Potřeba specifikovat počet shluků předem</i> • <i>Citlivá na odlehlé hodnoty („outliers“)</i>
TwoStep	<ul style="list-style-type: none"> • <i>Efektivně zvládá velké datové soubory</i> • <i>Pro spojité i kategorizované proměnné</i> • <i>Automaticky najde optimální počet shluků</i> 	<ul style="list-style-type: none"> • <i>Nepodporuje prázdná místa, chybějící hodnoty (vyloučí je)</i>

Zdroj:[29,37]

Před použitím vlastních modelovacích metod bylo ještě nutné převést spojitou proměnnou čistý měsíční příjem na kategoriální. Byly zvoleny 4 kategorie (nízký příjem, průměrný, nadprůměrný a vysoký příjem). Do těchto kategorií byla proměnná rozdělena z důvodu, že nelze porovnávat konkrétní hodnoty příjmu v roce 1997 a 2011. Výše příjmu se v těchto letech radikálně změnila a to díky růstu inflace a jiných ekonomických ukazatelů. V modelování tedy spíše šlo o to, zda existují nějaké rozdíly v příjmech u voličů, kteří preferují levicové resp. pravicové strany. Spojitá vstupní proměnná byla tedy rozdělena prostřednictvím asociačních pravidel do čtyř příjmových skupin a to podle zjištěných statistických hodnot minima, maxima a středních hodnot. Toto rozdělení je uvedeno v tabulkách 8 a 9.

Tabulka 8: Čistý měsíční příjem 1997

Cisty_mesicni_prijem	
Podmínka (IF)	Interval (THEN)
<=5000	<i>nízký</i>
>5000 and <=10000	<i>průměrný</i>
>10000 and <= 15000	<i>nadprůměrný</i>
>15000	<i>vysoký</i>

Zdroj: vlastní zpracování

Tabulka 9: Čistý měsíční příjem 2011

Cisty_mesicni_prijem	
Podmínka (IF)	Interval (THEN)
<=7000	<i>nízký</i>
>7000 and <=15000	<i>průměrný</i>
>15000 and <= 25000	<i>nadprůměrný</i>
>25000	<i>vysoký</i>

Zdroj: vlastní zpracování

Jak již bylo výše uvedeno, pro modelování byly využity dvě modelovací metody a to shluková analýza s algoritmem TwoStep a asociační pravidla. U asociačních pravidel byly při modelování využity dvě metody a to metoda Apriory a GRI. U obou asociačních metod bylo nutné nastavit hodnotu minimum antecedent support a minimum rule confidence. Nevhodnější hodnoty, při modelování s hodnotami z roku 1997, se jevily na úrovni 30%. Při těchto hodnotách bylo dosaženo nejlepších vypovídajících výsledků. Vygenerovaná pravidla obou metod jsou uvedeny v příloze C. Pro zjištění výstupní proměnné, typický volič pravice resp. levice v roce 1997, byla ještě využita shlukovací metoda a algoritmus TwoStep. U metody byl nastaven počet shluků, když je metoda schopna generovat automaticky počet shluků, jak je uvedeno v tabulce 7. Pokud metoda generovala počet shluků automaticky, nebyly výsledky příliš vhodné vzhledem k řešenému problému. Počet shluků tedy byl nastaven na 8. Výsledky této metody jsou uvedeny v příloze C. Výstupy z obou metod jsou uvedeny v následující tabulce 10. Získané výsledky s obou algoritmů asociačních pravidel jsou shodné, proto v tabulce 10 jsou uvedeny výsledky obou metod společně.

Na základě získaných výstupů z obou modelovacích metod (tabulka 10) lze říci, že typický volič, který preferuje pravicově zaměřené strany, dosahuje nižších měsíčních příjmů, vzhledem k ostatním osobám z datového souboru. S tím logicky souvisí fakt, že typickým voličem pravice je student. Studenti převážně nemají takové možnosti dosahovat na vyšší měsíční příjem. Naopak o levicově zaměřených osobách lze konstatovat, že jejich příjmy spadají do kategorie průměrných. A s tím lze opět logicky spojit sociální postavení zaměstnanec. Vysokoškoláci, kteří již pracují, mají možnost měsíčně vydělávat vyšší peněžní prostředky, než vysokoškoláci, kteří ještě studují. Dále o typických levicově zaměřených voličích lze říci, že jsou to převážně muži. Na základě výstupů z obou modelovacích metod, bohužel nelze konstatovat nic o pohlaví osob preferující pravicově zaměřené strany.

Tabulka 10: Výstupy z modelovacích technik pro rok 1997

		Asociační pravidla	TwoStep
Čistý měsíční příjem	PRAVICE	<i>nízký</i>	<i>nízký</i>
	LEVICE	<i>průměrný</i>	<i>průměrný</i>
Náзор na vývoj ČR	PRAVICE	<i>vyvíjí se správným směrem</i>	<i>vyvíjí se správným směrem</i>
	LEVICE	<i>nezjištěno</i>	<i>nezjištěno</i>
Pohlaví	PRAVICE	<i>žena</i>	<i>muž</i>
	LEVICE	<i>muž</i>	<i>muž</i>
Sociální postavení	PRAVICE	<i>student</i>	<i>student</i>
	LEVICE	<i>zaměstnanec</i>	<i>zaměstnanec</i>
Životní úroveň	PRAVICE	<i>spíše dobrá</i>	<i>spíše dobrá</i>
	LEVICE	<i>spíše dobrá</i>	<i>spíše dobrá</i>

Zdroj: vlastní zpracování

Pro porovnání a zjištění možných odlišností typického voliče z roku 1997 a 2011, bylo nutné provést modelování tedy ještě s daty získaných v roce 2011. Opět byly použity ty samé dvě modelovací metody. U shlukové analýzy byl použit algoritmus TwoStep a u asociačních pravidel algoritmus Apriori a GRI. Celkový postup modelování byl totožný s předchozím modelováním, proto jsou získané výsledky z obou metod uvedeny pouze v následující tabulce 11.

Tabulka 11: Výstupy z modelovacích technik pro rok 2011

		Asociační pravidla	TwoStep
Čistý měsíční příjem	PRAVICE	<i>nízký</i>	<i>nízký</i>
	LEVICE	<i>nízký</i>	<i>nízký</i>
Náзор na vývoj ČR	PRAVICE	<i>vyvíjí se špatným směrem</i>	<i>vyvíjí se špatným směrem</i>
	LEVICE	<i>vyvíjí se špatným směrem</i>	<i>vyvíjí se špatným směrem</i>
Pohlaví	PRAVICE	<i>žena</i>	<i>žena</i>
	LEVICE	<i>muž</i>	<i>muž</i>
Sociální postavení	PRAVICE	<i>student</i>	<i>student</i>
	LEVICE	<i>student</i>	<i>student</i>
Životní úroveň	PRAVICE	<i>velmi dobrá</i>	<i>velmi dobrá</i>
	LEVICE	<i>spíše dobrá</i>	<i>spíše dobrá</i>

Zdroj: vlastní zpracování

Z výše uvedené tabulky 11, ve které jsou uvedeny výstupy z modelovacích metod pro rok 2011, lze odvodit typického vysokoškolského voliče levice resp. pravice pro rok 2011. Z výše uvedených dat lze říci, že typickým voličem levice v roce 2011 je student, tedy osoba stále studující na vysoké škole, s tím samozřejmě souvisí i proměnná čistý měsíční příjem, který logicky odpovídá kategorii nízký. Důležitým postřehem je, že volič v roce 2011 má názor na vývoj ČR negativní, tedy že vývoj ČR směřuje špatným směrem. Tento fakt lze přisoudit špatné stávající ekonomické situaci, která se týká nejen ČR, ale i mnoha jiných států a to nejen v Evropské Unii. Pravicově zaměřené strany v roce 2011 spíše preferují studentky, naopak levicově zaměřené strany spíše preferují muži studující na vysoké škole. Obě skupiny vysokoškoláků žije, resp. v dětství žilo na velmi dobré resp. spíše dobré životní úrovni.

Komparace výstupů z modelovacích technik pro rok 1997 a 2011

Na základě získaných výstupů z modelování nad daty z roku 1997 a 2011 bylo provedeno porovnání a zjištění tak typického voliče pravice, resp. levice v roce 1997 a 2011. Zjištěné výstupy jsou zaznamenány v následující tabulce 12.

Tabulka 12: Komparace výstupů pro rok 1997 a 2011

		1997	2011
Čistý měsíční příjem	PRAVICE	<i>nízký</i>	<i>nízký</i>
	LEVICE	<i>průměrný</i>	<i>nízký</i>
Názor na vývoj ČR	PRAVICE	<i>vyvíjí se správným směrem</i>	<i>vyvíjí se špatným směrem</i>
	LEVICE	<i>vyvíjí se správným směrem</i>	<i>vyvíjí se špatným směrem</i>
Pohlaví	PRAVICE	<i>muž</i>	<i>žena</i>
	LEVICE	<i>nezjištěno</i>	<i>muž</i>
Sociální postavení	PRAVICE	<i>student</i>	<i>student</i>
	LEVICE	<i>zaměstnanec</i>	<i>student</i>
Životní úroveň	PRAVICE	<i>spíše dobrá</i>	<i>velmi dobrá</i>
	LEVICE	<i>spíše dobrá</i>	<i>spíše dobrá</i>

Zdroj: vlastní zpracování

Na základě výše uvedené tabulky lze konstatovat, že došlo k určitým změnám v charakteru typického voliče levice, resp. pravice, na základě zvolených vstupních

proměnných. Největší změna nastala v názoru na vývoj v České republice. Obě skupiny z cílového segmentu se většinou shodly na tom, že vývoj České republiky v roce 2011 směřuje špatným směrem. Změna tohoto názoru může být ovlivněna ne příliš dobrou ekonomickou situací v období 2011/2012, která se projevuje ve všech zemích Evropské Unie. Naopak cílový segment v roce 1997 zaujímá postoj k vývoji ČR spíše pozitivní, tedy že vývoj směřuje správným směrem. Osoby, které studují na vysoké škole, tedy z hlediska sociálního postavení studenti, preferují stále pravici. S tím souvisejí i získané výstupy z proměnné čistý měsíční příjem. Osoby preferující pravicové strany, jsou studenti, kteří spadají do nízké příjmové skupiny. Naopak u voličů preferujících levicové strany lze pozorovat určité změny. Levicové strany v roce 1997 preferovaly převážně zaměstnanci s průměrným měsíčním příjmem. V roce 2011 došlo ke změnám. Lze tedy říci, že studenti vysokých škol už ne tak převážně preferují pouze pravice, ale že v roce 2011 je poměr voličů levice, resp. pravice na vysokých školách vyrovnaný. Pohlaví není považováno za příliš silný faktor při vlivu na preference politických stran, což se také v rámci tohoto modelování prokázalo.

4.3. Model komparace hodnot a postojů vysokoškoláků v roce 1997 a 2011

Hodnoty

Hodnoty jsou považovány za základní stavební složku individuálního životního stylu. „*Hodnoty představují systém získaných dispozic člověka jednat nebo směřovat k cíli v souladu s žádoucností, kterou určují podmínky existence*“ [3, str. 203-204]. Hodnoty a hodnotová orientace predikuje chování a prožívání každého jedince v každodenních situacích v průběhu života. Na základě individuálních hodnot utváříme svůj život a výrazně jimi ovlivňujeme náš způsob života (životní styl). To, co je jedincem kladně hodnoceno, k čemu zaujímá pozitivní postoj a o co projevuje zájem, představuje pro něho určitou hodnotu. Hodnoty z psychologického hlediska možno pojímat jako výrazy vztahu mezi postojem a jeho objektem, jako osobní koncepce žádoucího, které se kromě uvědomělosti či alespoň uvědomitelnosti a formulovatelnosti vyznačují též značnou trvalostí a obecností. Z této obecnosti vyplývá, že specifické implikace uznávání určitých hodnot nenásledují ve specifických situacích s bezvýhradnou nutností. Hodnota nějakého objektu nebo činnosti je druhou stránkou jak jejich objektivních kvalit, uspokojujících příslušné potřeby, tak subjektivního vztahu k nim, vyjádřeného pozitivním postojem či zájmem. V rámci hodnotové orientace se jednotlivé hodnoty mohou realitně osamostatnit od původních potřeb a tento svůj

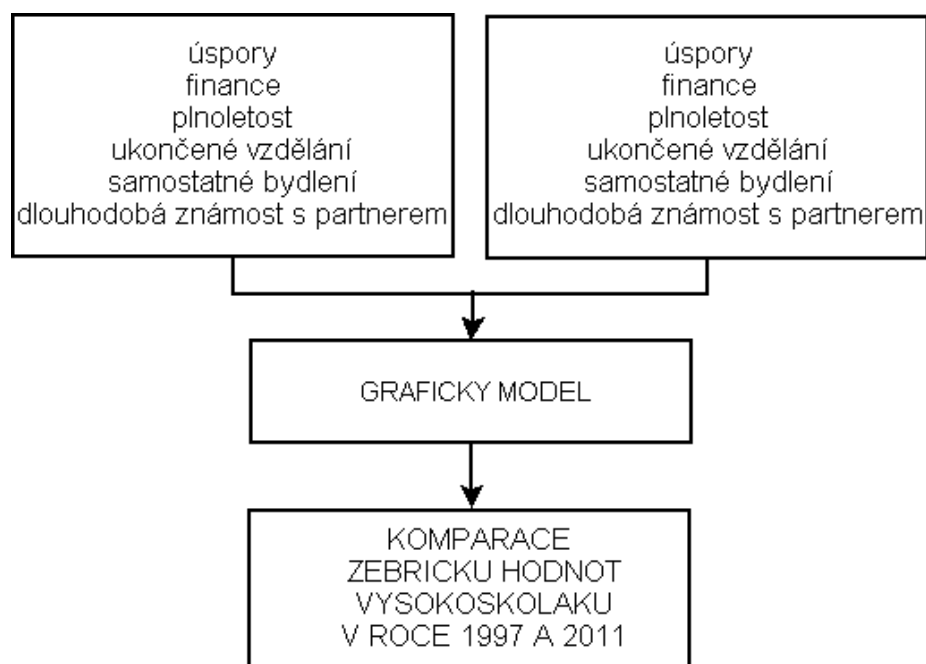
základ naopak zpětně modifikovat. Hodnoty potom představují nejen výraz uvědomění si nějakých potřeb, ale přímo osobní definici a hierarchizaci těchto potřeb. Hodnotová orientace je jako dominantní rámec spojující jednotlivé postoje zároveň vztažným rámcem veškerého dalšího hodnocení. [10,34]

Poznání hodnot je spjata s objektivním poznáním zákonitostí společenského vývoje. Hodnota je spojena se žádoucím chováním a cílovými stavy jedince nebo skupiny. Do hodnotové orientace se promítá vliv získaných zkušeností i hodnotové orientace sociální vrstvy, k níž jedinec náleží. Lze říci, že hodnoty jsou jakási pravidla (normy), které jedinci říkají, co je správné a co správné není. Hodnoty, které každý člověk vyznává, tvoří osobní a jedinečný osobní systém. Hodnoty můžeme dále považovat za nástroj sociálního přizpůsobení. Hodnoty představují individuální životní smysl a také jsou prostředkem pro řešení konfliktů v systému osobnosti (jedná se zejména o konflikt, který je způsoben rozhodováním). Hodnoty ovlivňují chování, integrují osobnost v kontextu důležitých životních cílů, zabraňují chaotickému jednání, ale jsou zároveň zdrojem konfliktů [10]

Cakirpaloglu [3] hovoří o takzvané triádě. Ta je nejdůležitější pro socializaci a hodnotovou stabilizaci jedince. Tvoří ji:

- Rodina
- Vzdělávací systém
- Zájmová nebo referenční skupina

Rodina je tedy jedním z nejdůležitějších faktorů pro hodnotovou stabilizaci jedince. Na základě těchto poznatků z tohoto zdroje byly vybrány dvě problematiky pořízení dětí a uzavírání manželství, která bezpochyby s problematikou rodina patří. Cílovým segmentem je opět skupina vysokoškoláků nebo vysokoškolsky vzdělaných lidí ve věkové kategorii 18-30 let. Následně bude sestaven žebříček hodnot cílového segmentu pro dvě vybrané problematiky a bude provedena komparace a posouzení vývoje ve změnách hodnot v letech 1997 a 2011. Žebříček hodnot obou problematik uvádí tabulky 13 a 14. Stanovené pořadí hodnot bylo realizováno v programovém prostředí Clementine. Jednotlivé hodnoty, které jsou uvedeny v žebříčku hodnot, byly vybrány jako jedny z nejdůležitějších vztahující se k těmto dvou problematikám, jak uvádí Sociologický ústav AV ČR, v.v.i. a to na základě ústavních analýz a výzkumů. Navržený obecný model vztahující se ke komparaci hodnot vysokoškoláků dané problematiky v roce 1997 a 2011 zobrazuje obrázek 10.



Obrázek 10: Obecný model - komparace hodnot vysokoškoláků v roce 1997 a 2011

Zdroj: vlastní zpracování

Žebříček hodnot si do jisté míry vytváří jedinec sám, nicméně je přitom ovlivňován systémem hodnot, které uznává společnost a ostatní lidé. [10]

4.3.1. Model komparace hodnot vysokoškoláků vztahující se k pořízení dětí v roce 1997 a 2011

Následující část práce je věnována hodnotám vysokoškoláků vztahující se k pořízení dětí v roce 1997 a 2011. Po zjištění daných hodnot je provedena komparace získaných výstupů. Následující tabulka 13 uvádí žebříček hodnot vztahující se k pořízení dětí vysokoškoláků v roce 1997 a v roce 2011.

Tabulka 13: Žebříček hodnot vztahující se k pořízení dětí pro rok 1997 a 2011

POŘÍZENÍ DĚTÍ	
1997	2011
PLNOLETOST	FINANCE
UKONČENÉ VZDĚLÁNÍ	DLOUHODOBÁ ZNÁMOST S PARTNEREM
SAMOSTATNÉ BYDLENÍ	PLNOLETOST
DLOUHODOBÁ ZNÁMOST S PARTNEREM	UKONČENÉ VZDĚLÁNÍ
FINANCE	SAMOSTATNÉ BYDLENÍ
ÚSPORY	ÚSPORY

Zdroj: vlastní zpracování

Při komparaci žebříčku hodnot vysokoškoláků vztahujících se k pořízení dětí jsou jasně viditelné rozdíly v roce 1997 a 2011. Zatím co v roce 1997 byl pro cílový segment nejdůležitější hodnotou pro pořízení dětí dovršení plnoletosti, v roce 2011 jsou pro cílový segment nejdůležitější hodnotou finance. Tento fakt, není pro současnou dobu a stávající ekonomickou situaci bezesporu nijak překvapující. Obě skupiny považují úspory za ne nijak významnou hodnotu pro pořízení dětí.

4.3.1. Model komparace hodnot vysokoškoláků vztahující se k uzavření manželství v roce 1997 a 2011

Níže uvedená tabulka 14 uvádí žebříček hodnot vysokoškoláků vztahující se k uzavření manželství. V tabulce jsou uvedeny hodnoty preferované vysokoškoláky v roce 1997 a v roce 2011.

Tabulka 14: Žebříček hodnot vztahující se k uzavření manželství pro rok 1997 a 2011

<i>UZA VŘENÍ MANŽELSTVÍ</i>	
<i>1997</i>	<i>2011</i>
PLNOLETOST	DLOUHODOBÁ ZNÁMOST S PARTNEREM
UKONČENÉ VZDĚLÁNÍ	PLNOLETOST
SAMOSTATNÉ BYDLENÍ	SAMOSTATNÉ BYDLENÍ
FINANCE	FINANCE
DLOUHODOBÁ ZNÁMOST S PARTNEREM	UKONČENÉ VZDĚLÁNÍ
ÚSPORY	ÚSPORY

Zdroj: vlastní zpracování

Stejně tak jako u žebříčku hodnot vztahující se k pořízení dětí, i zde jsou vidět patrné rozdíly v preferenci hodnot. Zatím co v roce 1997 cílový segment považoval kritérium ukončené vzdělání za poměrně důležité, v roce 2011 není toto kritérium pro uzavírání manželství pro cílový segment nijak důležité. Tento fakt lze podložit výsledky ze statistického úřadu, který uvádí, že od roku 2000 se razantně zvýšil počet uzavřených manželství mezi vysokoškolsky vzdělanými osobami, které jsou stále ještě studujícími. [36] Lze konstatovat, že kritérium plnoletost bylo v roce 1997 i v roce 2011 pro cílový segment poměrně důležitým kritériem. Cílový segment toto kritérium v roce 1997 dokonce uvedl na prvním místě žebříčku hodnot. Tento fakt lze opět podložit výsledky, získanými ze statistického úřadu. Ten uvádí, že průměrný sňatkový věk vysokoškoláků v roce 1997 je zhruba 28 let. Tato úroveň sňatkového věku se razantně liší od sňatkového věku „nevysokoškolských“ svateb. V tomto případě lze

mluvit o „doznívání“ obvyklého brzkého sňatkového věku, především tedy u „nevysokoškoláků“ [36]. U kritérií samostatné bydlení, finance a úspory nelze pozorovat žádné změny ve vývoji preferencí hodnot.

Postoj

Výrazná příbuznost mezi pojmy hodnoty a postoje vede k častým a nesprávným terminologickým záměnám a konfúzím. Postoje chápeme jako relativně dlouhodobé tendence k určitému způsobu myšlení, cítění a jednání, které se vztahují k lidem (včetně vlastní osoby), předmětům, činnostem i k dalším skutečnostem. V postojích je obsaženo subjektivní hodnocení těchto skutečností na základě jejich více či méně racionálního poznání, na základě jejich emocionálního působení na člověka i na základě vztahu k jeho cílům a záměrům. Podstatnou charakteristikou postojů je, že se ke svému objektu vztahují v určitém smyslu (pozitivně, neutrálně, negativně) a s určitou intenzitou (silně, slabě). Je důležité u obou pojmů odlišovat kognitivní, emoční a akční komponentu chování. Rozdíly mezi hodnotami a postoji lze vymezit poměrně přesně. Hodnoty zaujímají centrální místo v osobnosti, jsou částí osobnostní struktury, jsou odolnější ke změně v porovnání s postoji. Postoje nejsou tak abstraktní jako hodnoty. Postoje se vztahují ke konkrétním objektům či osobám, což nelze říci o hodnotách. Důležité je, že hodnoty jsou základem pro vznik konkrétních postojů. Obecně platí, že postoje dospělých jsou relativně stálější a tedy obtížněji měnitelné než u dětí. S tím také souvisí velmi častý případ z praxe a to cílevědomé formování, případně přetváření postojů. [3,15,21].

Potřeba

Potřeby a přístup k jejich uspokojení ovlivňují život jedince stejně jako hodnoty. Lze je považovat za jednu ze základních determinant lidského chování a prožívání. S pojmem potřeba se lze setkat jak v běžné řeči (potřeba=nedostatek), tak i v odborné terminologii. Potřebu můžeme definovat jako potenciál (připravenost organismu) reagovat určitým způsobem na danou situaci. Potřeba je v jeho pojetí hybná síla mozku, která nás nutí chovat se určitým způsobem s tím, aby bylo dosaženo cíle. Toto chování může být vzhledem k užitečnosti či příjemnosti orientováno cíleně. Chování je potřebami řízeno a odráží naši osobnost. Některé potřeby jsou dočasné a mění se, jiné vykazují časovou stálost a jsou hlouběji zvnitřněny v osobnosti člověka. Maslow rozlišil potřeby na nižší (např. jídlo, voda, bezpečí), vyšší (např. láska) a růstové (sebeaktualizace), které mají tvar pyramidy, kdy fyziologické potřeby tvoří základnu a potřeby sebeaktualizace její vrchol. Potřeby na úrovni 1

až 4 můžeme považovat za nedostatkové, potřeby na úrovni 5 definujeme jako existenční [5,10,11].

5. *Potřeby sebeaktualizace (stávat se lepším, být tím, čím mohu být).*

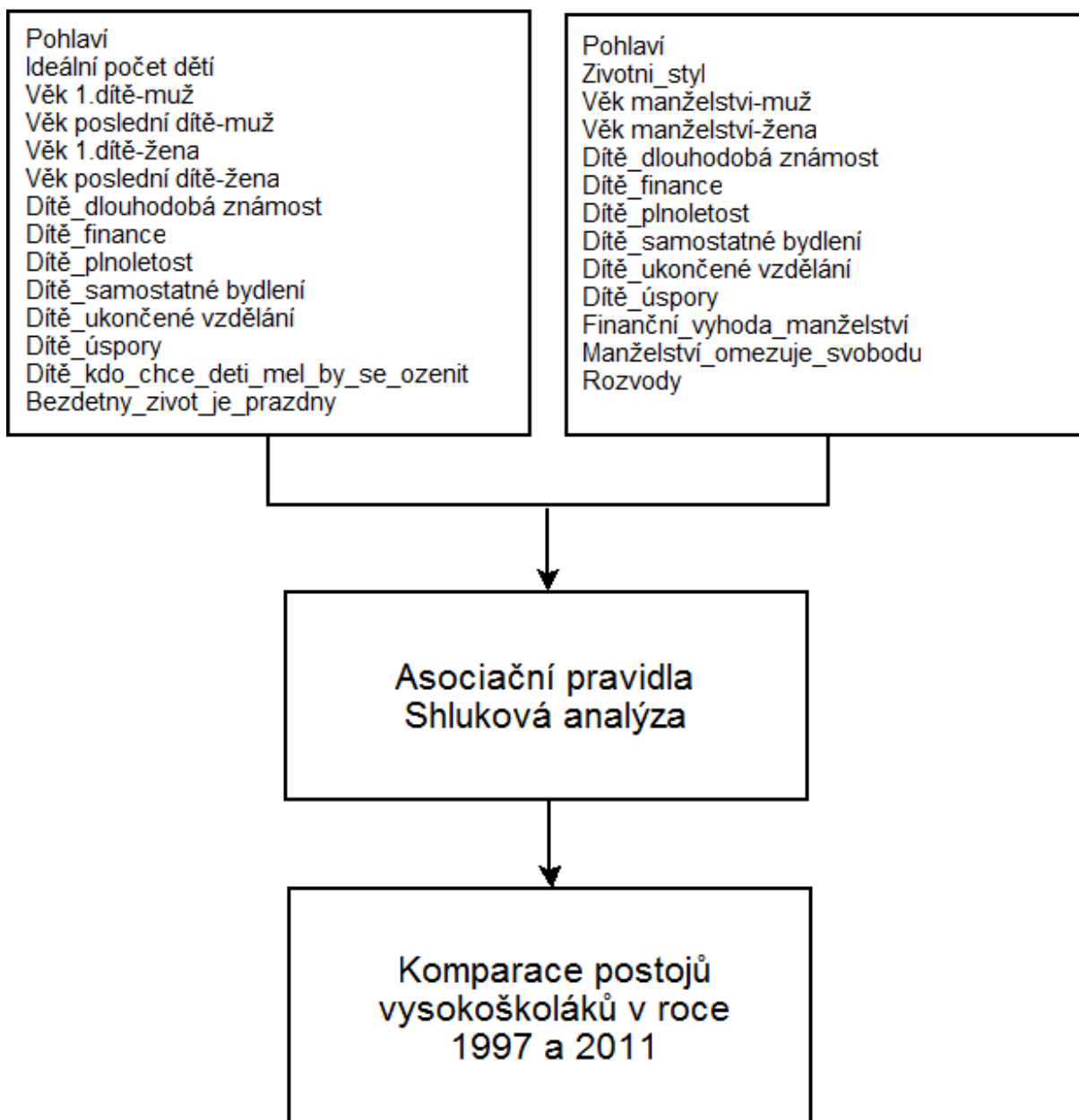
4. *Potřeby úcty a uznání/ prestiže (kladné sebehodnocení, pozitivní přijetí druhými lidmi).*

3. *Potřeby náležitosti a lásky (láska, něha, společenství, přátelství apod.).*

2. *Potřeby jistoty (svoboda, zbavení se strachu a úzkosti apod.).*

1. *Potřeby fyziologické (vzduch, jídlo, voda, sex apod.).*

V následující části je vytvořen model nad daty vztahující se k problematice výše uvedené. Tedy k pořízení dětí a uzavírání manželství. Pro zjištění preferujících hodnot a sestavení žebříčku hodnot vysokoškoláků vztahující se k daným problematikám bylo možné použít jednoduché grafy z programového prostředí Clementine. Pro zjištění postojů a pro následné porovnání těchto postojů, které zaujímali vysokoškoláci k daným problematikám v roce 1997 a 2011 jsou použity určité modelovací metody. Teoretické podklady k těmto použitým modelovacím metodám jsou uvedeny v kapitole „Vybrané modelovací metody“. Nejprve bude řešena problematika týkající se postojů vysokoškoláků v roce 1997 a 2011 k pořízení dětí. Následně bude obdobně vytvořen model a analýza tohoto modelu pro problematiku týkající se uzavírání manželství mezi vysokoškoláky v roce 1997 a v roce 2001. Po vytvoření obou modelů, bude provedena komparace postojů vysokoškoláků k dané problematice v obou uvedených letech. Obecný návrh modelu obou problematik zobrazuje obrázek 11. Z modelu jsou jasně patrné vstupy pro obě problematiky a modelovací techniky, které jsou využity. Model zobrazuje jediný výstup a to komparaci postojů vysokoškoláků v roce 1997 a 2011.



Obrázek 11: Obecný model - komparace postojů vysokoškoláků v roce 1997 a 2011

Zdroj: vlastní zpracování

4.3.2. Model komparace postojů vysokoškoláků k pořízení dětí v roce 1997 a 2011

V následující části je řešena problematika postoje vysokoškoláků k pořízení dětí v roce 1997 a 2011. Obecný návrh modelu zobrazuje obrázek 11. Pro zjištění podstaty problematiky byly využity dvě modelovací metody. Jedná se o shlukovou analýzu a asociační pravidla. U shlukové analýzy bylo pracováno s algoritmem TwoStep. Teoretický podklad této metody je

uveden v tabulce 7. Algoritmus TwoStep je schopen si vygenerovat počet shluků sám, ale podporuje i možnost vlastního nastavení počtu shluků. Obě tyto možnosti byly v práci využity a následně porovnány výstupy. Nejvhodnější pro další práci bylo vlastní nastavení shluků u obou případů, do kterých program sám objekty rozdělil. V tomto případě byly výstupy v přijatelné podobě. Při práci s asociačními pravidly byla převážně využita metoda Apriori. Výstupy z metody GRI v tomto případě nebyly příliš vypovídající vzhledem k řešenému problému, proto nebyly brány v úvahu. U metody Apriory byly stejně jako u prvního modelu nastavovány různé hodnoty antecedent support (podpora) a confidence (spolehlivost). Byly tedy vytvářeny různé modely, které byly navzájem porovnávány. Jako finální model byl v tomto případě zvolen model s minimální hodnotou antecedent support 40%. Jednalo se o model, který vykazoval nejvhodnější výsledky, vzhledem k ostatním. V příloze D jsou zobrazeny grafické výstupy vytvořených modelů. Nejsou zde uvedeny všechny výstupy z modelovacích metod, ale pouze ukázky výstupů z programového prostředí Clementine. Pro učinění závěrů byla samozřejmě selektována pouze některá pravidla. A to ta, která nabývala co možná nejvyšších hodnot spolehlivosti a podpory vzhledem k ostatním pravidlům. V následující tabulce 15, jsou zapsány výstupy získané z obou uvedených modelovacích metod.

Tabulka 15: Postoj vysokoškoláků k pořízení dětí v roce 1997 a 2011

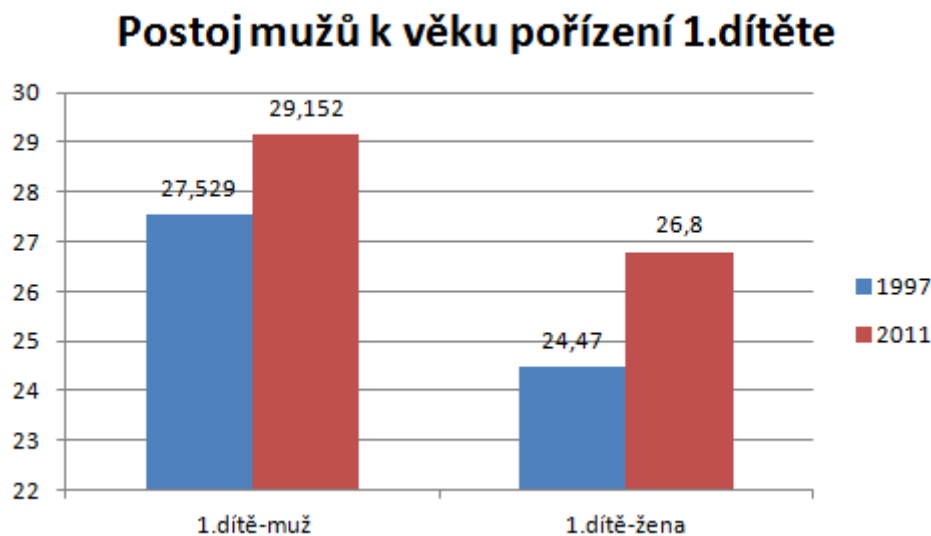
		POŘÍZENÍ DĚTÍ	
		1997	2011
Ideální počet dětí	MUŽ	2,23	2,109
	ŽENA	1,83	2,175
Věk-1.dítě-muž	MUŽ	27,529	29,152
	ŽENA	28,2	29,45
Věk-poslední dítě-muž	MUŽ	38,94	39,5
	ŽENA	39,6	41,45
Věk-1.dítě-žena	MUŽ	24,47	26,8
	ŽENA	23,6	26,75
Věk-poslední dítě-žena	MUŽ	34,17	36,6
	ŽENA	32,8	36,6
Dítě_dlouhodobá známost	MUŽ	<i>Nezjištěno</i>	<i>Velmi důležité</i>
	ŽENA	<i>Nezjištěno</i>	<i>Velmi důležité</i>
Dítě_finance	MUŽ	<i>Důležité</i>	<i>Velmi důležité</i>

		POŘÍZENÍ DĚTÍ	
		1997	2011
	ŽENA	<i>Důležité</i>	<i>Velmi důležité</i>
Dítě_plnoletost	MUŽ	<i>Velmi důležité</i>	<i>Velmi důležité</i>
	ŽENA	<i>Velmi důležité</i>	<i>Velmi důležité</i>
Dítě_samostatné bydlení	MUŽ	<i>Velmi důležité</i>	
	ŽENA	<i>Důležité</i>	
Dítě_ukončené_vzdělání	MUŽ	<i>Velmi důležité</i>	<i>Velmi důležité</i>
	ŽENA	<i>Velmi důležité</i>	<i>Velmi důležité</i>
Dítě_úspory	MUŽ	<i>Spíše nedůležitý</i>	<i>Neutrální</i>
	ŽENA	<i>Spíše nedůležitý</i>	<i>Neutrální</i>
Dítě_kdo_chce_deti_mel_by_se_oženit	MUŽ	<i>Spíše souhlasím</i>	<i>Spíše souhlasím</i>
	ŽENA	<i>Spíše souhlasím</i>	<i>Spíše souhlasím</i>
Bezdětný_život_je_prázdný	MUŽ	<i>Spíše nesouhlasím</i>	<i>Spíše nesouhlasím</i>
	ŽENA	<i>Spíše nesouhlasím</i>	<i>Spíše nesouhlasím</i>

Zdroj: vlastní zpracování

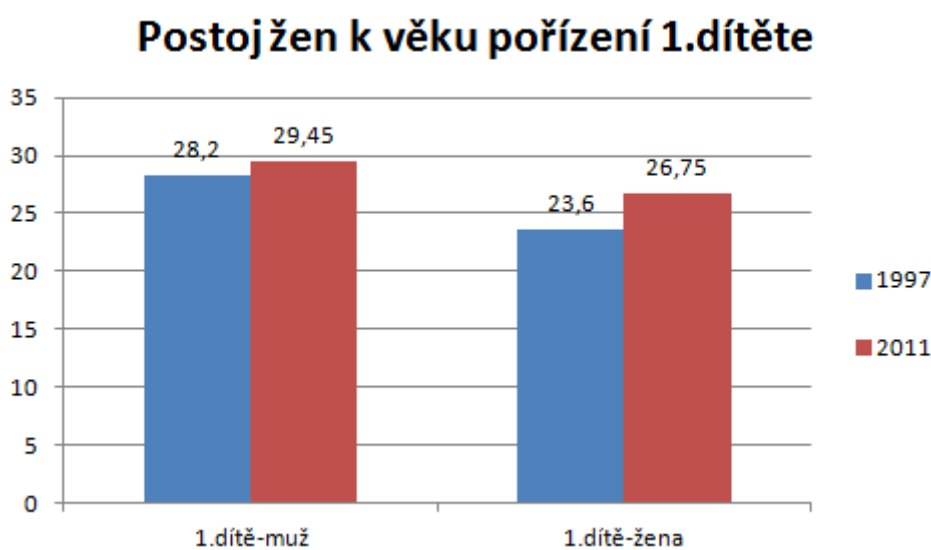
Údaje, které zobrazuje tabulka 15, představují výstupy ze dvou modelovacích metod, které byly již uvedeny výše. Jedná se o shlukovou analýzu a asociační pravidla. Cílem modelovacích technik bylo zjištění postojů vysokoškoláků k pořízení dětí v roce 1997 a 2011. Pokud provedeme analýzu modelu, lze vypočítat jisté změny v postojích vysokoškoláků, vztahujících se k pořízení dětí. Co se týče problematiky ideálního počtu dětí v rodině, obě skupiny vysokoškoláků, tedy z roku 1997 i z roku 2011, považují stále 2 děti za ideální počet dětí v rodině. Pokud porovnáme postoj vysokoškoláků k věkové hranici, kdy by si měli muži, resp. ženy pořizovat první, resp. poslední dítě, je zde patrný nárůst věkové hranice. To, že za poslední roky došlo k nárůstu věkové hranice pro pořizování dětí obecně, tedy bez rozlišení vzdělání, dokládají i podklady získané ze statistického úřadu. Podle údajů ze statistického úřadu byl průměrný věk matky, v okamžiku narození jejího prvního dítěte v roce 1996 26,1 let. V roce 2010 už tato věková hranice u žen vzrostla na 29,6 let. [20] Tyto údaje vypovídají o všech ženách, tedy bez rozlišení podle jejich vzdělání. Na základě získaných dat z výše uvedeného modelu lze potvrdit i tento rostoucí trend u skupiny vysokoškoláků. Vysokoškoláci v roce 2011 uvádějí ve všech případech vyšší věkovou hranici, než byla

uvedena v roce 1997. Tyto výstupy z modelu byly zjištěny pomocí shlukové analýzy. Grafický výstup výsledků zobrazují dva níže uvedené grafy 2 a 3.



Graf 2: Postoj mužů k věku pořízení 1. dítěte

Zdroj: vlastní zpracování



Graf 3: Postoj žen k věku pořízení 1. dítěte

Zdroj: vlastní zpracování

Na výše uvedených grafech 2 a 3 je graficky zobrazena věková hranice, vyjadřující postoj mužů, resp. žen k věku pořízení prvního dítěte. Jak je výše uvedeno, z grafu je možné vyčíst jistý rostoucí trend věkové hranice, jak u žen, tak u mužů ve vztahu k pořízení prvního dítěte. Pokud porovnáme postoj žen k věkové hranici pořízení prvního dítěte a skutečné věkové hodnoty, kdy si ženy pořizovaly první děti, jsou zde patrné rozdíly v obou letech.

V roce 1997 ženy preferují ideální věkovou hranici 23,6 pro pořízení dítěte, ale hodnoty získané ze statistického úřadu, uvádějí, že v roce 1996/1997 si ženy pořizovaly první dítě v průměru až v 26,1 letech. Takový to věkový rozdíl lze zpozorovat i v roce 2011. Údaje ze statistického úřadu vyjadřují průměrný věk pořízení prvního dítěte u žen na 29,6 let, zatímco ženy zaujímají postoj k věku pořízení prvního dítěte o něco jiný. Jako ideální věk pořízení prvního dítěte preferují hranici 26,75.

Prostřednictvím asociačních pravidel, bylo zjišťováno, jaký postoj zaujímají ženy, resp. muži k pořízení dětí a to na základě vstupních proměnných, které jsou uvedeny v tabulce 15. Tyto vstupní proměnné vycházejí z dotazníku, který je v příloze A. Podle analýz ze sociologického ústavu AV ČR, v.v.i, se jedná o proměnné, které nejvíce ovlivňují postoj člověka při pořízení dětí. Z tohoto důvodu, byly také tyto proměnné zvoleny jako vstupy do modelu. U některých proměnných nebylo možno zjistit jistý výstup, proto jsou některá pole v tabulce prázdná. V následující analýze modelu budou tedy porovnány pouze ty proměnné, u kterých jsou zjištěné výstupy jak pro rok 1997, tak pro rok 2011. Postoje vysokoškoláků k problému vycházejí ze škály, která je uvedena v dotazníku (příloha A). Jednotlivé proměnné hodnotili na stupnici důležitosti a to od 1 do 5, přičemž 1 bylo nejméně důležitá a 5 nejvíce důležitá. 5 tedy znamená velmi důležité, 4 důležité, 3 neutrální, 2 spíše nedůležité a 1 absolutně nedůležité. Poslední dvě proměnné vycházejí ze stupnice 1 - určitě souhlasím, 2 - spíše souhlasím, 3 - spíše nesouhlasím a 4 - určitě nesouhlasím.

Jak tabulka uvádí, je patrné, že faktor finance, jsou pro skupinu vysokoškoláků v roce 2011 velmi důležité pro pořízení dětí. Oproti tomu, skupina vysokoškoláků v roce 1997 faktor finance neuvádí, jako jeden z nejdůležitějších pro pořízení dítěte. Pro tuto skupinu je více než finance důležité dosažení plnoletosti a ukončené vzdělání pro pořízení dětí. O problematice úspor lze říci, že ani jedna skupina vysokoškoláků nepřikládá příliš velkou důležitost tomuto faktoru při pořízení dětí. Obě skupiny zaujímají stejný postoj k problematice pořízení dětí a manželství. I přes dnešní vysoký a stále narůstající počet narozených dětí mimo manželství, jak uvádí statistický úřad [20], obě skupiny se jednoznačně shodli na tom, že uzavření manželství před narozením dítěte je nezbytné. Negativní postoj zaujímají k problematice života bez dětí, jako prázdného života. Obě skupiny se opět shodly na tom, že život bez dětí rozhodně nemůže být považován za prázdný.

Na závěr lze tedy o tomto modelu říci, že postoj vysokoškoláků k pořízení dětí se od roku 1997 až do současnosti příliš nijak výrazně nezměnil. Došlo pouze k výraznější změně postoje k faktorům finance a úspory. Současní vysokoškoláci považují mít dostatek financí za

velmi důležité pro pořízení dětí a stejně tak i o úsporách lze říci, že pro současné vysokoškoláky jsou o něco důležitější než pro vysokoškoláky v roce 1997 v situaci pořízení dítěte. Výrazná změna byla zjištěna u postoje k věkové hranici pořízení prvního, resp. posledního dítěte u ženy, resp. muže. Celková věková hranice se posunula směrem výše, průměrně o dva roky u všech případů. Tento vzrůstající fakt potvrdily i údaje získané ze statistického úřadu [20].

4.3.3. Model komparace postojů vysokoškoláků k uzavření manželství v roce 1997 a 2011

V následující části je řešena problematika postoje vysokoškoláků k uzavření manželství v roce 1997 a v roce 2011. Obecný návrh modelu zobrazuje obrázek 11. Pro zjištění podstaty problematiky byly využity opět dvě modelovací metody. Jedná se o shlukovou analýzu a asociační pravidla. U shlukové analýzy bylo pracováno s algoritmem TwoStep. Teoretický podklad této metody je uveden v tabulce 7. Algoritmus TwoStep je schopen si vygenerovat počet shluků sám, ale podporuje i možnost vlastního nastavení počtu shluků. Obě tyto možnosti byly v práci využity a následně porovnány výstupy. Při práci s asociačními pravidly byly v tomto případě využity obě možnosti, tedy Apriori i GRI.

U metody Apriory byly stejně jako u prvního modelu nastavovány různé hodnoty antecedent support a confidence. Byly tedy vytvářeny různé modely, které byly navzájem porovnávány. Jako finální model bylo zvoleno vždy takové nastavení, které vykazovalo nejvhodnější požadované výstupy. Pro následnou komparaci výsledků, byly vždy vybírány pravidla, jejichž hodnoty support (podpora) a confidence (spolehlivost) byly co možná nejvyšší vzhledem k ostatním pravidlům. Jednalo se o model, který vykazoval nejvhodnější výsledky, vzhledem k ostatním. V příloze E jsou zobrazeny grafické výstupy vytvořených modelů. V následující tabulce 16, jsou zapsány výstupy získané z obou uvedených modelovacích metod.

Tabulka 16: Postoj vysokoškolačů k uzavření manželství v roce 1997 a 2011

		POŘÍZENÍ DĚTÍ	
		1997	2011
Věk-1.manželství-muž	MUŽ	27,225	28,216
	ŽENA	27,137	28,38
Věk-1.manželství-žena	MUŽ	23,626	26,216
	ŽENA	24,118	26,4
Dítě_dlouhodobá známost	MUŽ	<i>Nezjištěno</i>	<i>Velmi důležité</i>
	ŽENA	<i>Nezjištěno</i>	<i>Velmi důležité</i>
Dítě_finance	MUŽ	<i>Důležité</i>	<i>Velmi důležité</i>
	ŽENA	<i>Důležité</i>	<i>Velmi důležité</i>
Dítě_plnoletost	MUŽ	<i>Velmi důležité</i>	<i>Důležité</i>
	ŽENA	<i>Velmi důležité</i>	<i>Velmi důležité</i>
Dítě_samostatné bydlení	MUŽ	<i>Velmi důležité</i>	<i>Důležité</i>
	ŽENA	<i>Důležité</i>	<i>Velmi důležité</i>
Dítě_ukončené_vzdělání	MUŽ	<i>Důležité</i>	<i>Neutrální</i>
	ŽENA	<i>Velmi důležité</i>	<i>Neutrální</i>
Dítě_úspory	MUŽ	<i>Neutrální</i>	<i>Spíše nedůležité</i>
	ŽENA	<i>Neutrální</i>	<i>Spíše nedůležité</i>
Životní styl	MUŽ	3	3
	ŽENA	3	3
Rozvody	MUŽ	1	1
	ŽENA	1	3

Zdroj: vlastní zpracování

Údaje, které zobrazuje tabulka 16, představují výstupy ze dvou modelovacích metod, které byly již uvedeny výše. Jedná se o shlukovou analýzu a asociační pravidla. Cílem modelovacích technik bylo zjištění postojů vysokoškolačů k uzavření manželství v roce 1997 a v roce 2011. Pokud provedeme analýzu modelu, lze vypočítat jisté změny v postojích vysokoškolačů, vztahujících se k pořízení dětí. Proměnné, u kterých nebylo možné zjistit všechny údaje v obou letech, nebudou brány při analýze a komparaci v úvahu.

Poslední dvě proměnné se týkají postojů, jaké vysokoškolači zaujímají k rozvodům a životnímu stylu vysokoškolačů, jaký preferují před uzavřením manželství. Z výstupů, které jsou uvedeny v tabulce 16, je patrné, že muži i ženy vysokoškolači v obou letech zaujímají

stále stejný postoj ke způsobu života s partnerem/partnerkou před uzavřením svatby. Jednoznačně se shodují, že je nevhodnější žít s partnerem/partnerkou na „zkoušku“ a později zavřít manželství. Lze tedy konstatovat, že postoj k životnímu stylu se nezměnil. Co se týče problematiky rozvodů, jisté změny tu lze zaznamenat. Jedná se hlavně o změny u žen v postoji k této problematice. Zatímco muži stále pohlížejí na rozvod tak, že by to měl být jednoznačně jednoduchou záležitostí, náhled žen na rozvody se od roku 1997 do roku 2011 změnil. V dnešní době ženy pohlížejí na řešení problematiky rozvodů velmi negativně. K rozvodům zaujímají takový postoj, že rozvody by neměly být rozhodně snadnou záležitostí pro žádné manželství. Pokud se dále zaměříme na problematiku věkové hranice uzavírání prvního manželství u žen a mužů vysokoškoláků, i zde jsou patrné rozdíly. Stejně jako u výše uvedeného modelu, který se zabýval komparací postojů vysokoškoláků k pořízení dětí, tak i zde je zaznamenána zvýšená věková hranice uzavírání manželství. Tento fakt lze opět podložit výsledky analýz ze statistického úřadu, který uvádí, že věková hranice uzavírání manželství u všech lidí, tedy bez ohledu na vzdělání, vzrostal v průměru o 2,7 let u mužů a o 2,8 let u žen od roku 1997 do roku 2011 [31]. Pokud tedy vezmeme výsledky z tabulky 16, která se týká vysokoškoláků lze říci, postoj k věkové hranici týkající se uzavírání prvních manželství u žen vzrostl zhruba o 1,92 let a u mužů zhruba o 0,991 let. Z těchto údajů tedy vyplývá, že trend stále se zvyšující hranice uzavírání manželství se potvrdil i u skupiny vysokoškoláků a to jak u žen, tak i u mužů. Jedná se tedy o trend, který se v současné době dotýká všech skupin obyvatelstva, bez ohledu na pohlaví či ukončené vzdělání.

U zbývajících proměnných (samostatné bydlení, finance, plnoletost, ukončené vzdělání, úspory a dlouhodobá známost), které se vztahují k postoji k uzavírání manželství, byla opět využita hodnotící škála, která je uvedena v dotazníku (příloha A). Jednotlivé proměnné hodnotili respondenti na stupnici důležitosti a to od 1 do 5, přičemž 1 byla nejméně důležitá a 5 nejvíce důležitá. 5 tedy znamená velmi důležité, 4 důležité, 3 neutrální, 2 spíše nedůležité a 1 absolutně nedůležité. Jednoznačná změna je u proměnné finance. Zatímco v roce 1997 nebyl tento faktor při uzavírání manželství pro vysokoškoláky jedním z nejdůležitějších, v roce 2011 tento faktor považují vysokoškoláci za velmi důležitý. Tento fakt lze již vypožorovat v tabulce 14, ve které jsou uvedeny preference hodnot vysokoškoláků. Jednoznačně staví proměnnou finance na první místo. U proměnné plnoletost nedošlo k příliš velkým změnám v postojích vysokoškoláků. Stále tento faktor považují za velmi důležitý, aby mohli uzavřít manželství. Snad jen ženy v současnosti zaujímají mírnější postoj, myslí si tedy, že to není tak důležité pro uzavření manželství. Došlo však k poměrně velkým změnám v postojích k ukončenému vzdělání a uzavření prvního manželství.

Vysokoškoláci v roce 1997 považují tento fakt za poměrně důležitý. Naopak v dnešní době není přikládán k tomuto faktoru příliš velký důraz. Obě skupiny vysokoškoláků se shodují, že vlastnit určitý objem úspor není nijak zvlášť důležitým faktorem pro uzavření manželství.

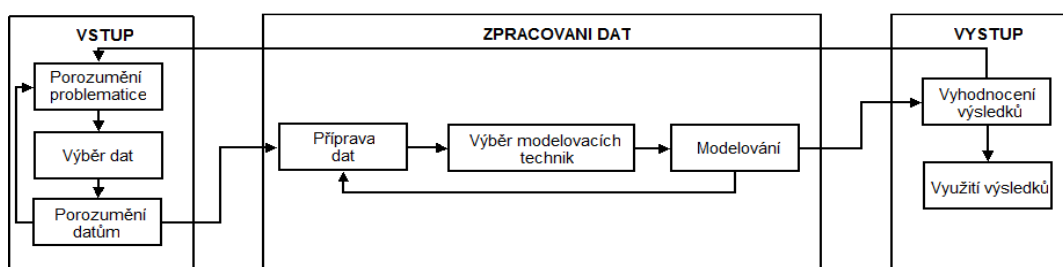
Na závěr lze tedy o tomto modelu říci, že v postojích vysokoškoláků k uzavírání prvních manželství od roku 1997 až do současnosti došlo k určitým změnám. K poměrně značné změně došlo v postoji k financím. Dnešní vysokoškoláci tento faktor považují za jeden z nejdůležitějších a to jak v uzavírání manželství, tak i v pořízení dětí, jak je patrné z předešlého modelu. Další výraznější změna nastala v postoji k ukončenému vzdělání. Dnešní vysokoškoláci nepovažují tento faktor za příliš důležitý. Výrazná změna však byla zjištěna u postojů k věkové hranici uzavírání prvního manželství u ženy, resp. muže. Celková věková hranice se posunula směrem výše, průměrně o dva roky u žen a průměrně o jeden rok u mužů. Tento vzrůstající fakt potvrdily i údaje získané ze statistického úřadu [31].

Data byla zpracována v programovém prostředí Clementine. Zde byla data připravena před samotnou realizací vybraných modelovacích metod. Přípravou dat se rozumí odstranění odlehlých hodnot, vyřešení problematiky chybějících hodnot nebo chybných hodnot. Veškerá tato problematika je podrobněji popsána v kapitole příprava dat. Výsledný stream z programu je uveden v příloze F.

Data byla, jak je výše uvedeno, zpracovávána v programu Clementine. Zde byly také realizovány vybrané modelovací metody. Clementine je aplikace pro obchodní analýzy. Program je velmi rozšířený a oblíbený u data miningových pracovníků a analytiků po celém světě. Jedná se o komplexní nástroj pro data mining („dolování dat“). Oblíbený je zejména proto, že umožňuje celý data miningový proces od vložení dat až po zobrazení skórovacích výsledků. Velké možnosti dále umožňuje zejména ve velmi efektivním grafickém rozhraní a neschopnosti systematizovat sestavován a ukládání analytických postupů pro případné pozdější použití. Program umožňuje dále uživatelům snadnou přípravu číselných, textových i webových dat. Dále lze také rychle vytvářet různé modely, které navzájem lze porovnávat a vyhodnocovat. Clementine vytváří proudy datového procesu, v nichž se nachází jednotlivé kroky toku data miningových úloh. Velkou oblíbenost Clementine nese díky tomu, že analytik může v jakémkoliv kroku přidat svou odbornou znalost, což je klíčem k úspěšnému data miningu. Proces data miningu je v programu realizován velmi rychle a efektivně. Tím tedy šetří čas a náklady na projekt. Poslední verzí je Clementine 11. Tato verze je označována za verzi „posilující obrannou analýzu“. Poptávka po této verzi se zvýšila hlavně ve dvou sektorech: ve zdravotnictví a v národní bezpečnosti. Clementine 11 se zaměřuje přímo na to,

aby vyhověla právě potřebám těchto dvou sektorů, ale dělá to tak, aby současně nabídla spoustu aplikací také firmám na ostatních trzích.[17,29]

Při práci bylo po celou dobu modelování vycházeno z metodiky CRISP-DM⁹. Obecný návrh metodiky zobrazuje obrázek 12. Jedná se o metodiku, která vznikla v rámci Evropského výzkumného projektu. Hlavním cílem bylo navrhnout univerzální postup použitelný v různých komerčních aplikacích. Podstatou tedy bylo navrhnout jakýsi standardní model procesu dobývání znalostí v oblasti databází. Vytvořená metodika umožňuje rychlejší, efektivnější, spolehlivější a méně nákladnější řešení rozsáhlých úloh dobývání znalostí. Problematika metodiky CRISP-DM je podrobně rozebrána v příloze G.



Obrázek 12: Návrh obecného modelu

Zdroj: upraveno podle [17]

Při práci bylo nutné rozlišovat typy dat. Velmi důležité to bylo zejména z toho důvodu, že každá modelovací technika vyžaduje jiné typy vstupních dat a ne každá modelovací technika zvládne pracovat se všemi typy dat. Než přijde na samotnou část modelování, je tedy nezbytně nutné porozumět tomu, jakým způsobem jsou data klasifikována.

Při klasifikaci dat rozlišujeme dvě třídy:

- Kvalitativní
- Kvantitativní

Rozlišujeme čtyři typy kvantitativních dat:

- Nominální data
- Ordinální data
- Intervalová data
- Spojitá data

⁹ CRISP-DM - Cross-Industry Standard Process for Data Mining

Podrobně rozebraná problematika klasifikace dat je uvedena v příloze G. V této části jsou podrobně definovány jednotlivé třídy dat a následně i typy kvantitativních dat. Tato problematika je doplněna o grafické zobrazení typů dat. [17]

V programu Clementine, ve kterém byly aplikovány na data modelovací techniky, lze pracovat s několika datovými typy dat. Typy dat, která byla využita v této práci, jsou uvedeny v datovém slovníku. V programu Clementine tedy rozlišujeme:

- Rozsah (Range)
- Diskrétní (Discrete)
- Příznak (Flag)
- Množina (Set)
- Uspořádané množina (Order Set)
- Bez typu (Typeless)

Podrobně rozebraná problematika typu dat v programovém prostředí Clementine je opět uvedena v příloze G.

ZÁVĚR

Cílem této diplomové práce bylo definování vybraných základních pojmů v oblasti sociální politiky, návrh modelů vybrané oblasti sociální politiky a analýza modelů. V první kapitole práce jsem podrobněji definovala pojmy z oblasti sociální politiky. Bližší pozornost jsem věnovala problematice sociologii, která je považována za jednu ze sociálních věd. Dále jsem zde podrobněji řešila problematiku sociální politiky obecně a problematiku sociální politiky zaměřující se konkrétně na Evropskou unii. Krátce jsem také definovala problematiku Evropské sociální charty a velmi často v současnosti diskutované termíny sociální inženýrství a sociální podnikání. Při návrhu modelů jsem využila určité modelovací metody, které jsou teoreticky podloženy v druhé části této práce. Jedná se o shlukovou analýzu, rozhodovací stromy a asociační pravidla. Data, která jsem zpracovávala, pocházejí ze dvou různých zdrojů. První soubor dat jsem získala z databáze sociologického ústavu AV ČR, v.v.i. Z tohoto zdroje jsem také převzala dotazník, na základě něhož, jsem provedla vlastní dotazníkové šetření a získala druhý datový soubor.

V další části této práce se věnuji popisu a analýze vstupních dat týkající se obou datových souborů. Před zahájením modelování bylo nutné data určitým způsobem upravit. Jednalo se především o vyřešení problematiky chybějících dat či chybných dat. V případě, že u záznamu od určitého respondenta bylo velké množství chybějících dat, tento záznam z datového souboru jsem zcela vyloučila. V případě, že se jednalo pouze o jeden nebo malý počet chybějících dat, využila jsem určité metody nahrazení chybějících dat. V této práci jsem využila substituci jedné hodnoty. Po zjištění, že data se vykazují jako nerovnoměrně rozložená, jsem pro nahrazení chybějících hodnot využila medián.

V práci jsem vytvořila a navrhla dva konkrétní modely. První model se zabývá problematikou politických stran, druhý model se věnuje komparaci postojů a hodnot vysokoškoláků k problematice pořízení dětí a uzavírání manželství v roce 1997 a 2011. Před samotnou realizací modelu a využití vybraných modelovacích metod jsem provedla analýzu dat. K provedení analýzy a zjištění výsledků prvního modelu jsem využila pouze grafický nástroj v programu Clementine. Výsledkem této analýzy jsem zjistila, jaká část respondentů se v roce 1997 a v roce 2011, považuje za pravicově orientované a jaká část za levicově orientované. Následně jsem tyto výstupy z obou let porovнала. Výsledek této analýzy je zobrazen v tabulce 3 a grafu 1. Výsledky jasně ukazují na to, v roce 1997 větší část vysokoškoláků preferovala „levicové“ strany, naopak v roce 2011 vysokoškoláci spíše preferují „pravicové“ strany. Tuto problematiku jsem řešila z důvodu, že podíl

vysokoškoláků, účastnících se ve volbách, je v porovnání s ostatními skupinami lidí, celkem malý. Zajímavé tedy bylo zjištění, po porovnání výsledků s celkovými výsledky v jednotlivých letech, že skupina vysokoškoláků nemá nijak výrazně odlišný názor na politickou orientaci než ostatní skupiny občanů. Výsledky voleb v roce 2010, které zobrazuje tabulka 6, ukazují na velmi malou převahu „pravice“ nad „levicí“. Zhruba 65% vysokoškoláků v roce 2011 preferuje „pravici“ a zhruba 35% „levici“. Zbylé osoby se charakterizují jako nestranní. Po analýze dat následovalo vytvoření první modelu. Návrh obecného modelu je zobrazen na obrázku 9. Výstupem z prvního modelu byla proměnná, která charakterizovala typického voliče „levice“ a „pravice“ v roce 1997 a 2011. Při modelování jsem využila dvě modelovací metody, shlukovou analýzu a asociační pravidla. Při práci s asociačními pravidly jsem využila dvě metody a to APRIORY a GRI. Jako vstupní atributy pro část pravidla „antecedent“ (předpoklad) byly vybrány atributy názor na vývoj ČR, životní úroveň respondenta v dětství, pohlaví a sociální postavení, věková kategorie a čistý příjem. Pro pole „consequent“ (závěr) byl vybrán atribut politická orientace (pravice, levice, střed). Některé výše uvedené atributy jsem nejprve převedla z dat spojitých na data kategorická. Obě dvě metody vygenerovaly určitá pravidla. Krom asociačních pravidel jsem využila i shlukovou analýzu, ve které jsem pracovala s algoritmem TwoStep, vzhledem ke své charakteristice, že je schopen pracovat i s kategoriálními daty, jak je uvedeno v tabulce 7. Výstupy z obou technik jsem na závěr porovnála a vyvodila z nich závěry. Výstupy z obou modelovacích technik zobrazují tabulky 10 a 11. Grafické výstupy získané z programu Clementine jsou uvedeny v příloze C. Jednotlivé charakteristiky typického voliče „levice“ resp. voliče „pravice“ v roce 1997, resp. v roce 2011 jsou zaznamenány v tabulce 12 a výsledky jsou následně slovně porovnány.

Druhý model se týká komparace postojů a hodnot vysokoškoláků v roce 1997 a v roce 2011 k pořízení dětí a uzavírání manželství. Před samotným vytvořením modelu jsem opět provedla analýzu dat. Cílem této analýzy dat bylo sestavení žebříčku hodnot vysokoškoláků v roce 1997 a v roce 2011 vztahující se k pořízení dětí a uzavření manželství. Žebříčky hodnot jsou uvedeny v tabulkách 13 a 14. Jednotlivé faktory, ze kterých respondenti sestavovali svůj žebříček hodnot k dané problematice, představují finance, úspory, samostatné bydlení, plnoletost, ukončené vzdělání a dlouhodobá známost s partnerem, jak je zobrazeno na obrázku číslo 10. Výsledky jsem získala pomocí grafického nástroje v programu Clementine. Ze získaných výsledků, lze konstatovat poměrně značné změny v preferencích hodnot vysokoškoláků v roce 1997 a 2011 vztahující se k pořízení dětí a uzavírání manželství. Poměrně očekávaným výsledkem byla jednoznačná preference faktoru finance pro vysokoškoláky

v roce 2011. Tento fakt lze odůvodnit nedávnou „ekonomickou krizí“ a současnými „úspornými opatřeními“. Konkrétní zhodnocení výstupů jsem opět uvedla u dané problematiky v poslední kapitole. Po provedení analýzy dat a sestavení žebříčku hodnot, jsem dále provedla samotné modelování pomocí vybraných modelovacích metod. Návrh obecného modelu této řešené problematiky zobrazuje obrázek číslo 11. Při modelování jsem použila shlukovou analýzu a asociační pravidla. Výstupem z druhého modelu je proměnná, která vyjadřuje postoje vysokoškoláků k pořízení dětí a uzavírání manželství. Konkrétní definice termínu postoj a hodnota a vysvětlení jejich odlišností, jsem podrobněji definovala v poslední kapitole u daného problému. Při práci s asociačními pravidly jsem opět využila dva modely a to APRIORY a GRI. Jako vstupní atribut pro část pravidla „antecedent“ (předpoklad) byly v prvním případě vybrány atributy ideální počet dětí, dlouhodobá známost, finance, plnoletost, samostatné bydlení, ukončené vzdělání, úspory, kdo chce mít děti, měl by se oženit, bezdětný život je prázdný. Pro pole „consequent“ (závěr) byl vybrán atribut pohlaví. U druhého modelu, vztahující se k postoji vysokoškoláků v obou letech k uzavírání manželství jsem vybrala jako vstupní atributy životní styl, dlouhodobá známost, finance, plnoletost, samostatné bydlení, ukončené vzdělání, úspory, rozvody, finanční výhoda manželství a manželství omezuje svobodu. Pro pole „consequent“ (závěr) byl opět vybrán atribut pohlaví. Dále jsem ještě použila shlukovou analýzu. V prvním případě jsem použila proměnné věk-první dítě-muž (žena), věk-poslední dítě-muž (žena) a pohlaví. V druhém případě jsem použila proměnné věk-první manželství-muž (žena). Výstupy z jednotlivých metod u jednotlivých modelů byly vždy zaznamenány do tabulky. Jedná se o tabulky číslo 15 a 16 a grafy 2 a 3. Následně jsem provedla komparaci výsledků získaných od respondentů v roce 1997 a v roce 2011. Konkrétní zhodnocení výsledků a modelu je uvedeno v poslední kapitole, vždy za každým modelem. Grafické výstupy z modelovacích technik z programového prostředí Clementine jsou uvedeny vždy v příloze. Jedná se o přílohy D a E.

Cílem této práce bylo získání dat z oblasti sociální politiky a tato data pomocí data-miningových metod zpracovat. Data jsem získala, provedla jsem analýzu dat a pomocí shlukové analýzy a asociačních pravidel provedla návrh a tvorbu modelů. Při práci jsem ve velké míře využívala znalosti získané díky studiu na Fakultě ekonomicko-správní Univerzity Pardubice a troufám si říci, že získané zkušenosti budu moci využít později v praxi. Cíl této práce byl tímto splněn.

POUŽITÁ LITERATURA

- [1] BERKA, Petr. *Dobývání znalostí z databází*. 1. vyd. Praha : Praha : Academia, 2003. 366 s. ISBN 80-200-1062-9
- [2] BRDEK, Miroslav; JÍROVÁ, Hana. *Sociální politika v zemích EU a ČR*. 1.vyd. Praha : CODEX Bohemia, s.r.o., 1998. 392 s. ISBN 80-85963-71-X.
- [3] CAKIRPALOGLU, Panajotis. *Psychologie hodnot*. Olomouc : Olomouc:Votobie, 2004. 427 s. ISBN 80-7220-195-6.
- [4] DISMAN, Milan. *Jak se vyrábí sociologická znalost : příručka pod uživatele*. 3. vyd. Praha : Karolinum, 2009. 374 s. ISBN 80-246-0139-7.
- [5] DRAPELA, Viktor J. *Přehled teorií osobnosti*. Praha: Portál, 1998. 175 s.
- [6] FEDINA, Martin. *Metafory v politickém diskurzu Václava Klause* [online]. *E-polis.cz*, 19. září 2011. [cit. 2011-11-14]. Dostupné z WWW: <<http://www.e-polis.cz/politicke-teorie/659-metafory-v-politickem-diskurzu-vaclava-klause.html>>. ISSN 1801-1438.
- [7] HINDLS, Richard, et al. *Analýza dat v manažerském rozhodování*. 1. vyd. Praha : Grada, 1999. 360 s. ISBN 80-7169-255-7.
- [8] KUBANOVÁ, Jana. *Statistické metody pro ekonomickou a technickou praxi*. 3. vyd. Bratislava : Bratislava : Stasis, 2008. 247 s. ISBN 978-80-85659-47-4.
- [9] KUNŠTÁT, Daniel. SOÚ AV,ČR v.v.i. *Sociální a demografický profil voličského zázemí českých politických stran*. Praha.
- [10] MANDYS, Jan. *Životní styl lidí bez domova*. Ostrava, 2011. 224 s. Dizertační práce. Ostravská univerzita v Ostravě.
- [11] MASLOW, Abraham H. *Motivation and personality*. Harper & Row, Publishers, 1987. 293 s.
- [12] MELOUN, Milan; MILITKÝ, Jiří. *Statistická analýza experimentálních dat*. 2.vyd. Praha : Academia, 2004. 953 s. ISBN 80-200-1254-0.
- [13] NĚMEC, Bohumil, et al. *Ottův slovník naučný-Nové doby : Dodatky k velkému Ottovu slovníku naučnému*. Praha : "NOVINA" , 1940. sociologie, s. 70-71.
- [14] NOVOTNÁ, Eliška. *Základy sociologie*. 1. vyd. Praha : GradaPublishing, a.s., 2008. 192 s. ISBN 978-80-247-2396-9.
- [15] PAULÍK, Karel. *Psychologie osobnosti pro praxi*. Ostrava, 2007. Učební text pro distanční studium. Ostravská univerzita v Ostravě.

- [16] PAYNE, Jan. *Kvalita života a zdraví*. 1. vyd. Praha: TRITON, 2005. ISBN 80-7254-657-0.
- [17] PETR, Pavel. *Data Mining : Díl I.* 2.vyd. Pardubice : Univerzita Pardubice, 2008. 144 s. ISBN 978-80-7395-098-9.
- [18] PHILLIPS, David. *Quality of Life : Concept, policy and practice*. 1.vyd. London New York : London New York : Routledge, 2006. 247 s. ISBN 978-0-415-32355-0.
- [19] PLECITÁ-VACHOVÁ, Klára. Levice a pravice. *Revuepolitika: Politicko-společenská revue Centra pro studium demokracie a kultury* [online]. 7/2001, 2001-01-20 [cit. 2012-02-29]. ISSN 1803-8468. Dostupné z: <http://www.revuepolitika.cz/clanky/1111/levice-a-pravice>.
- [20] Porodnost a plodnost 1996-2010: Počet a struktura narozených. In: *Český statistický úřad* [online]. 2007-09-27 [cit. 2012-03-19]. Dostupné z: [http://www.czso.cz/csu/2011edicniplan.nsf/t/5D002BEE05/\\$File/400811a1.pdf](http://www.czso.cz/csu/2011edicniplan.nsf/t/5D002BEE05/$File/400811a1.pdf)
- [21] RAPLEY, Mark. *Quality of Life Research : A Critical Introduction*. 1.vyd. London : London : Sage Publications, 2003. 251 s. ISBN 978-0-7619-5457-6.
- [22] Rozhodovací stromy [online]. 2007 [cit. 2001-11-28]. Rozhodovací stromy. Dostupné z WWW: <http://datamining.xf.cz/view.php?cisloclanku=2002102802>.
- [23] RYDZI, Daniel; RAUCH, Jan. Aplikace asociačních pravidel ve společnosti Zinest s.r.o.. *Systémová integrace* [online]. 2008, 4/2008, [cit. 2012-02-08]. Dostupný z WWW: <http://www.cssi.cz/cssi/aplikace-asociacnich-pravidel-ve-spolecnosti-zinest-s-r-o>.
- [24] ŘEHAN, Vladimír, ÇAKIRPALOGLU, Panajotis. Sociální status a hodnotová orientace mladé generace. In *Československá psychologie*. 2000, r. XLIV, č. 3, s. 202 – 215.
- [25] ŘEHÁK, Jan; ŘEHÁKOVÁ, Blanka. *Analýza kategorizovaných dat v sociologii*. 1.vyd. Praha : Academia, 1986. 400 s. ISBN 21-060-85.
- [26] ŘEHÁKOVÁ, Blanka. SOCIOLOGICKÝ ÚSTAV AV ČR. *Sociální trendy: Volební chování různých skupin voličů v předčasných parlamentních volbách v roce 1998*. Praha: a, 1998. ISBN 80-85950-52-9.
- [27] Sociální podnikání přináší veřejný prospěch. *Corpor AID Magazin*. 2008, 3, s. 3. Dostupný také z WWW: http://www.arsczech.org/files/publikace_prosinec_DESIGN.pdf.
- [28] Sociologický ústav Akademie věd ČR,v.v.i. *Český sociálně vědní datový archiv* [online]. 1997 [cit. 2011-10-28]. Dostupné z WWW: <http://nesstar.soc.cas.cz/webview/>.

- [29] SPSS. *SPSS Inc. Clementine 10.1 Desktop User's Guide*. 2006.
- [30] TESTYSOFTWARE : Clementine 10 posiluje obrannou analýzu. *COMPUTERWORLD*. 2006, 19, s. 21. Dostupný také z WWW: <http://www.spss.cz/files/marketing/computer_world2006.pdf>.
- [31] Územní diference sňatků a rozvodů: Průměrný věk snoubenců. In: *Český statistický úřad* [online]. 2010-09-27 [cit. 2012-03-21]. Dostupné z: [http://http://www.czso.cz/csu/2004edicniplan.nsf/t/69003B7057/\\$File/snatky.pdf](http://http://www.czso.cz/csu/2004edicniplan.nsf/t/69003B7057/$File/snatky.pdf)
- [32] VAĎUROVÁ, Helena; MÜHLPACHR, Pavel. *Kvalita života : Teoretická a metodologická východiska*. 1. vyd. Masarykova univerzita v Brně, 2005. 132 s. ISBN 80-210-3754-7.
- [33] VÍTEK, Jan. *Data Mining Solutions* [online]. 28.10.2002 [cit. 2011-10-31]. Metodologie CRISP-DM. Dostupné z WWW: <<http://datamining.xf.cz/view.php?cislocianku=2002102807>>.
- [34] VLÁČIL, Jiří. K pojetí hodnot v psychologii a sociologii. In *Československá psychologie*. 1977, r. XXL, č. 1, s. 72 – 85.
- [35] Volby.cz. *Český statistický úřad* [online]. 2010, 2012-02-03 [cit. 2012-02-29]. Dostupné z: <http://www.volby.cz/>
- [36] Vysokoškoláci z demografického pohledu 1991-2006: Sňatky. In: *Český statistický úřad* [online]. 2007-09-27 [cit. 2012-03-15]. Dostupné z: [http://www.czso.cz/csu/2007edicniplan.nsf/t/6100309267/\\$File/403607k04.pdf](http://www.czso.cz/csu/2007edicniplan.nsf/t/6100309267/$File/403607k04.pdf)
- [37] *Web Data Mining*. [online]. 2010 [cit. 2012-03-07]. Dostupné z WWW: <http://books.google.cz/books?id=6Mh50Uaq6AIC&pg=PA124&lpg=PA124&dq=strengths+K-means&source=bl&ots=NuvTCNIono&sig=PVfM_hy9_LX8kO5wSSgSZ5R8pDE&hl=cs&ei=RsHVS9mbMseLOMa3vJAO&sa=X&oi=book_result&ct=result&resnum=3&ved=0CA4Q6AEwAg#v=onepage&q=strengths%20K-means&f=false>.
- [38] ŽIŽKOVÁ, Jana, et al. *Základy sociální politiky*. 1.vyd. Praha : Vysoká škola ekonomická v Praze, 1993. 157 s. ISBN 80-7079-975-7.

SEZNAM PŘÍLOH

Příloha A	Sociologický dotazník	<i>Zdroj: upraveno podle [26]</i>
Příloha B	Příprava dat-nahrazení chybějících hodnot	
Příloha C	Grafické zobrazení výstupů modelů – politická orientace	
Příloha D	Grafické zobrazení výstupů modelů – postoje vysokoškoláků k pořízení dětí	
Příloha E	Grafické zobrazení výstupů model – postoje vysokoškoláků k uzavírání manželství	
Příloha F	Výsledný stream z programu Clementine	
Příloha G	Metodik CRISP a typy dat	

Příloha A - Sociologický dotazník

- 1. V politice se často používají pojmy „pravice“ a „levice“. Kam byste se sám(a) zařadil(a), kam svými názory patříte?**
 - levice
 - střed
 - pravice
- 2. Myslíte si, že se situace v České republice vyvíjí všeobecně správným anebo nesprávným směrem?**
 - vyvíjí se správným směrem
 - vyvíjí se špatným směrem
 - nesměřuje nikam
- 3. Vezmete-li v úvahu všechny okolnosti svého života, řekl(a) byste, že jste celkově:**
 - velmi spokojený
 - spíše spokojený
 - spíše nespokojený
 - velmi nespokojený
- 4. V jaké rodině jste prožil převážnou část dětství a dospívání? Dětstvím se rozumí období do Vašich 12 let, dospíváním období od 12 do 18 let. (dětství označte DE, dospívání-DO)**
 - v úplné rodině s oběma vlastními rodiči
 - v nové úplné rodině vzniklé po rozvodu/úmrtní jednoho z rodičů
 - v nesezdaném soužití vlastních rodičů (vztah druh – družka)
 - v neúplné rodině pouze s matkou
 - v neúplné rodině pouze s otcem
 - u příbuzných
 - jinde
- 5. Jak byste hodnotil(a) životní úroveň rodiny, v níž jste v období dětství a dospívání vyrůstal(a)? (dětství označte jako DE, dospívání-DO)**
 - velmi dobrá
 - spíše dobrá
 - spíše špatná
 - velmi špatná
- 6. S kolika sourozenci (vlastními i nevlastními) jste vyrůstal(a) ve společné domácnosti?**
 - v dětství _____
 - v dospívání _____
- 7. Jakým způsobem většinou řešíte situace, které přinášejí život?**
 - určitě rázně - rozhoduji se okamžitě, bez váhání
 - spíše rázně - rozhoduji se po krátké rozvaze
 - spíše umírněným - rozhoduji se až po delší úvaze
 - určitě umírněným - dlouho váhám, než přijmu rozhodnutí

8. **Z následujících možností vyberte jednu, která nejlépe vystihuje Vaši povahu a náuru:**

- každou změnu vítám a sám(a) věci rád(a) měním
- změny mi nevadí, ale nevyhledávám je
- změny nemám příliš rád(a), ale dokážu se s nimi vyrovnat
- změny mi vadí, těžko se s nimi vyrovnávám a nerad(a) věci měním

9. **Nyní, pomocí pětibodové stupnice ohodnoťte, jak jsou následující věci pro Vás osobně důležité. 1 vyjadřuje nejmenší důležitost, 5 naopak největší.**

- A/ mít co nejvyšší vzdělání a kvalifikaci _____
- B/ mít čas na vlastní zájmy a koníčky _____
- C/ mít zajímavou a užitečnou práci _____
- D/ žít pro svou rodinu a děti _____
- E/ mít úspěch a uznání _____
- F/ mít čas na přátele _____
- G/ mít hodně peněz a dobře si žít _____

10. **Každý člověk má určité představy o tom, jaké vlastnosti by měl mít jeho životní partner. Řekněte nám, jakou důležitost Vy sám(a) přikládáte u životního partnera následujícím charakteristikám. (1-nejmenší důležitost;5-největší důležitost)**

- A/ vysokoškolské vzdělání _____
- B/ dobré finanční a hmotné zabezpečení _____
- C/ hezký vzhled a pěkné vystupování _____
- D/ sportovní založení _____
- E/ zodpovědnost a poctivost _____
- F/ dobrý vztah k dětem _____
- G/ inteligence _____
- H/ tolerantnost _____
- I/ přizpůsobivost _____

11. **Představte si, že byste se rozhodl(a) oženit (vdát). Hrály by při Vašem rozhodování o uzavření manželství důležitou roli následující důvody?**

(-určitě ano-UA; spíše ano-SA; spíše ne-SN; určitě ne-UN)

- A/ nesouhlas rodičů s nesezdaným soužitím _____
- B/ snaha, aby se děti narodily v manželství _____
- C/ možnost seberealizace _____
- D/ touha mít bezpečné zázemí _____
- E/ jistota bezpečného pravidelného sexuálního života _____
- F/ finanční zajištění _____
- G/ jasné majetkové vztahy _____

12. **Kterou z následujících možností považujete pro svůj život za nejlepší?**

- žít sám/sama bez stálého partnera
- žít se stálým partnerem/partnerkou bez sňatku
- žít s partnerem/partnerkou "na zkoušku" a později uzavřít sňatek

-uzavřít s partnerem/partnerkou sňatek bez předchozího života na zkoušku

13. Když vezmete v úvahu všechny okolnosti, jaký je Váš názor na nejvhodnější věk pro:

-uzavření prvního manželství u muže _____ ženy _____

- narození prvního dítěte u muže _____ ženy _____

-narození posledního dítěte u muže _____ ženy _____

-ideální počet dětí v rodině _____

14. Za jak důležité považujete následující okolnosti pro to, aby mladý člověk mohl uzavřít sňatek? (1-nejméně důležité;5-nejvíce důležité)

A/ dovršení plnoletosti _____

B/ ukončení vzdělání _____

C/ ekonomická samostatnost, výdělečná činnost _____

D/ dostatečný finanční příjem _____

E/ úspory kolem 100 tisíc Kč _____

F/ možnost samostatného bydlení _____

G/ dostatečně dlouhá známost s partnerem _____

15. Budoucí plány ohledně dětí?

-počet dětí zatím neplánuju

-nechci mít děti

-plánuji děti

16. Počet plánovaných dětí? _____

17. Za jak důležité považujete následující okolnosti pro to, aby se člověk odpovědně rozhodl pro narození dítěte? (1-nejméně důležité;5-nejvíce)

A/ dovršení plnoletosti _____

B/ ukončení vzdělání _____

C/ dostatečný finanční příjem _____

D/ úspory kolem 100 tisíc Kč _____

E/ možnost samostatného bydlení _____

F/ dostatečně dlouhá známost s partnerem _____

18. Nyní prosím uveďte, jak souhlasíte s následujícími výroky:

(určitě souhlasím-US;spíše souhlasím-SS;spíše nesouhlasím-SN;určitě nesouhlasím-UN)

-hlavní výhodou manželství je finanční jistota _____

-hlavním důvodem uzavírání sňatků je mít děti _____

- lidé, kteří chtějí mít děti, by měli uzavřít sňatek _____

- sledovat, jak děti rostou, je největší radost v životě _____

- mít děti znamená příliš velké omezení svobody rodičů _____

- lidé, kteří nikdy neměli děti, vedou prázdný život _____

- manželství vždy omezuje osobní svobodu muže a ženy _____

-v nejbližších letech se v naší zemi zlepší životní úroveň většiny obyvatel _____

- v nejbližších letech se v naší zemi zvýší dostupnost bydlení pro mladé lidi _____

-každý pracovitý člověk v aktivním věku si dnes v naší zemi může vydělat hodně peněz _____

- nezaměstnanost se v naší zemi nebude v nejbližších letech podstatně zvyšovat _____

19. Jaké je podle Vás v současné době oficiálně stanovené životní minimum pro 3-člennou rodinu, v níž žijí rodiče s jedním dítětem ve věku do 6 let? _____

20. Který z následujících výroků nejlépe vystihuje Váš názor na rozvody?

-rozvod by měl být jednoduchou záležitostí.

-jsou-li v rodině děti, rozvod by měl být nesnadný, aby se udrželo rodinné prostředí pro děti.

-rozvod by neměl být snadnou záležitostí pro žádné manželství.

-rozvod by měl být zakázán.

21. Pohlaví respondenta:

1. muž

2. žena

22. Jaké je Vaše sociální postavení?

-student, učeň

-nezaměstnaný(á)

-zaměstnanec

-družstevník

-živnostník, soukr. farmář

-podnikatel

-ostatní

23. Jaké je Vaše nejvyšší ukončené školní vzdělání?

-základní bez vyučení

-vyučen(a) bez maturity

-maturita

-vysoká škola (VOŠ)

24. Studujete v současné době? ANO NE

25. Jaká je Vaše finanční situace? Odhadněte

-svůj průměrný čistý měsíční příjem Kč

-své úspory Kč

26. Velikost obce, ve které máte trvalý pobyt:

-do 2000 obyvatel

-2000-9 999 obyvatel

-10 000-49 999 obyvatel

-50 000-99 999 obyvatel

-velkoměsto nad 100 000 obyvatel

27. Vyberte území, ve kterém žijete:

-Praha

- Střední Čechy
- Jižní Čechy
- Západní Čechy
- Severní Čechy
- Východní Čechy
- Jižní Morava
- Severní Morava

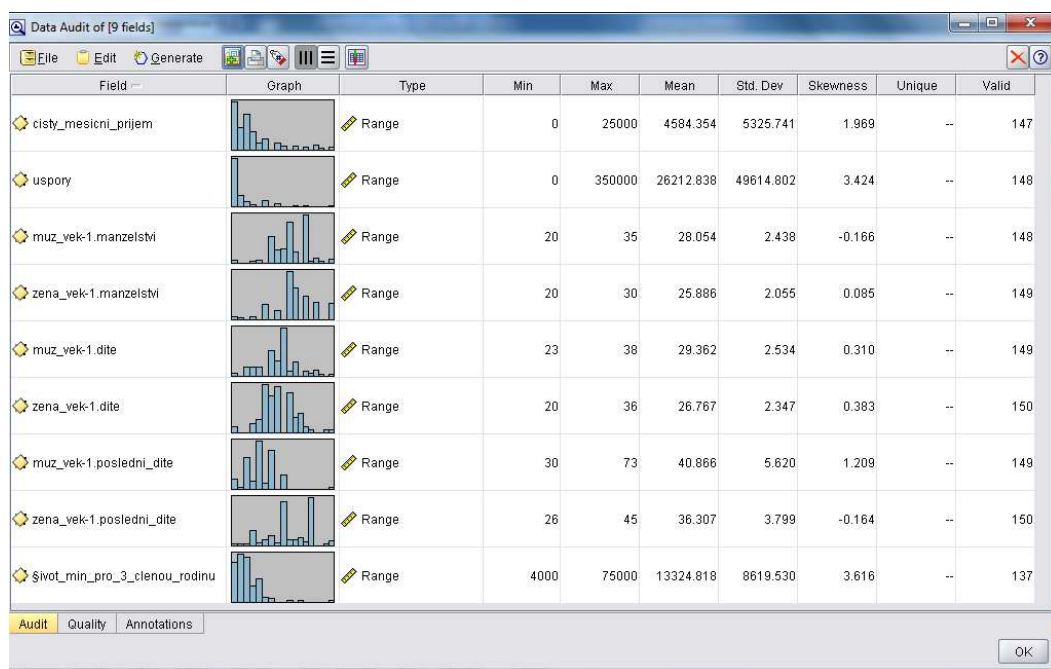
28. Věková kategorie, do které spadáte:

- 18-20 let
- 20-23 let
- 24-26 let
- 27-29 let

29. Čistý osobní měsíční příjem:

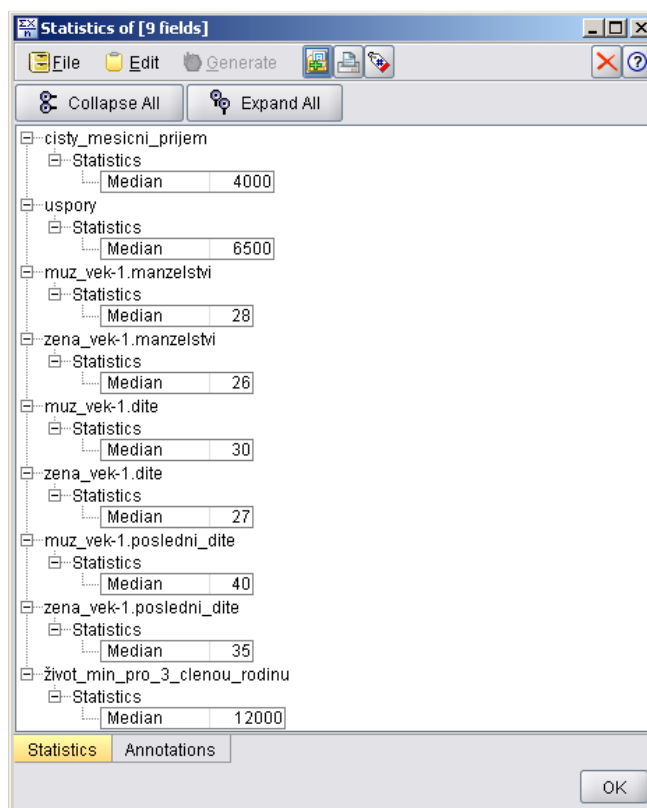
- méně než 3 999 Kč
- 4000- 5 499 Kč
- 5 500- 7 499 Kč
- 7 500- 9 999 Kč
- 10 000 a více

Příloha B - Příprava dat-nahrazení chybějících hodnot



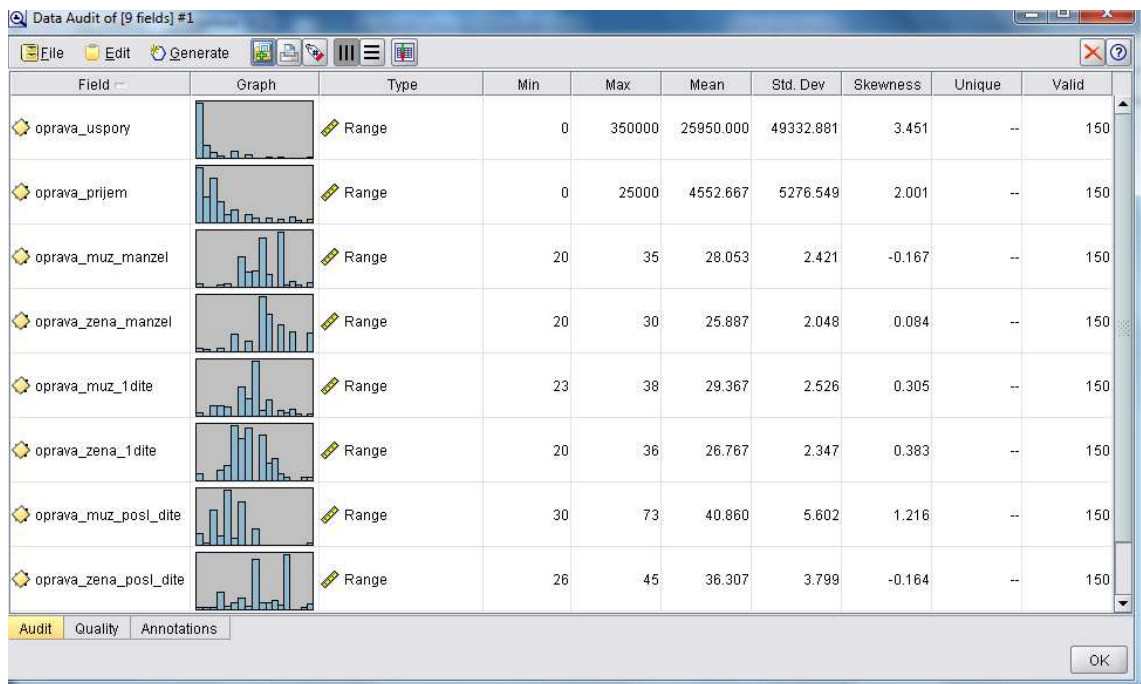
Obrázek 13: Zobrazení nerovnoměrného rozložení dat

Zdroj: vlastní zpracování



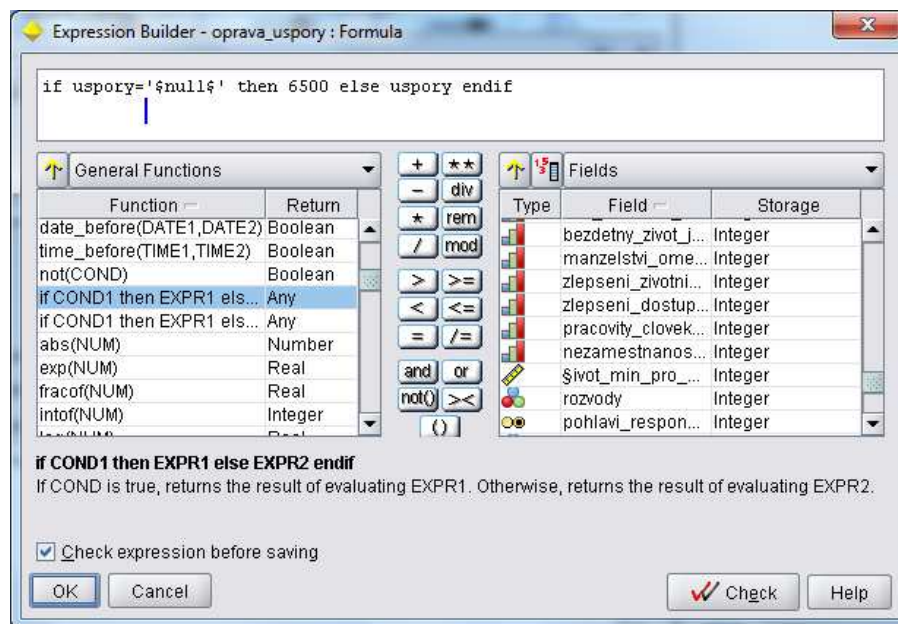
Obrázek 14: Hodnot mediánu pro nahrazení chybějících dat

Zdroj: vlastní zpracování



Obrázek 15: Statistika nahrazených chybějících hodnot

Zdroj: vlastní zpracování



Obrázek 16: Postup nahrazování chybějících hodnot

Zdroj: vlastní zpracování

Příloha C - Grafické zobrazení výstupů modelů – politická orientace

Consequent	Antecedent	Support %	Confidence %
polit_orientace = pravice	vyvoj_v_CR = 1	30,769	67,857
polit_orientace = pravice	zivotni_uroven_detstvi = 2		
polit_orientace = pravice	vyvoj_v_CR = 1	49,451	66,667
polit_orientace = pravice	soc_postaveni = student	35,165	62,5
polit_orientace = pravice	soc_postaveni = student		
polit_orientace = pravice	prijem_kategorie = nizky	31,868	58,621
polit_orientace = pravice	vyvoj_v_CR = 1		
polit_orientace = pravice	soc_postaveni = zamestna...	30,769	57,143
polit_orientace = pravice	prijem_kategorie = nizky	46,154	54,762
polit_orientace = levice	prijem_kategorie = prumerny		
polit_orientace = levice	soc_postaveni = zamestna...	34,066	51,613
polit_orientace = levice	soc_postaveni = zamestna...		
polit_orientace = levice	zivotni_uroven_detstvi = 2	40,659	48,649
polit_orientace = pravice	pohlavi = muz	43,956	47,5
polit_orientace = levice	prijem_kategorie = prumerny	37,363	47,059
polit_orientace = pravice	pohlavi = zena	56,044	45,098
polit_orientace = pravice	pohlavi = zena		
polit_orientace = pravice	soc_postaveni = zamestna...	37,363	44,118
polit_orientace = pravice	zivotni_uroven_detstvi = 2	64,835	44,068
polit_orientace = pravice	pohlavi = zena		
polit_orientace = pravice	zivotni_uroven_detstvi = 2	35,165	43,75
polit_orientace = levice	vyvoj_v_CR = 3	30,769	42,857
polit_orientace = levice	pohlavi = muz	43,956	42,5
polit_orientace = levice	soc_postaveni = zamestna...	64,835	42,373
polit_orientace = pravice	prijem_kategorie = prumerny	37,363	38,235
polit_orientace = levice	zivotni_uroven_detstvi = 2	64,835	37,288
polit_orientace = pravice	soc_postaveni = zamestna...	64,835	37,288
polit_orientace = pravice	soc_postaveni = zamestna...		
polit_orientace = pravice	zivotni_uroven_detstvi = 2	40,659	32,432

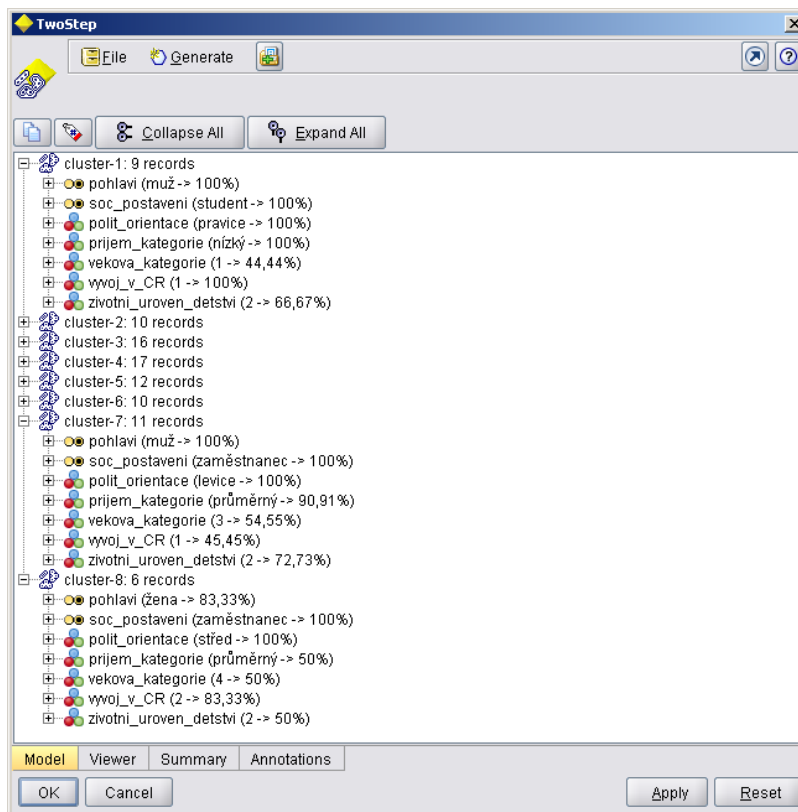
Obrázek 17: Vygenerovaná pravidla pro rok 1997 - Apriori – model 1

Zdroj: vlastní zpracování

Consequent	Antecedent	Support %	Confidence %
polit_orientace = prav...	vyvoj_v_CR = 1	49,45	66,67
polit_orientace = levice	zivotni_uroven_detstv...	40,66	48,65
	soc_postaveni = zam...		
polit_orientace = levice	soc_postaveni = zam...	64,84	42,37
polit_orientace = levice	pohlavi = muz	43,96	42,5

Obrázek 18: Vygenerovaná pravidla pro rok 1997 - metoda GRI

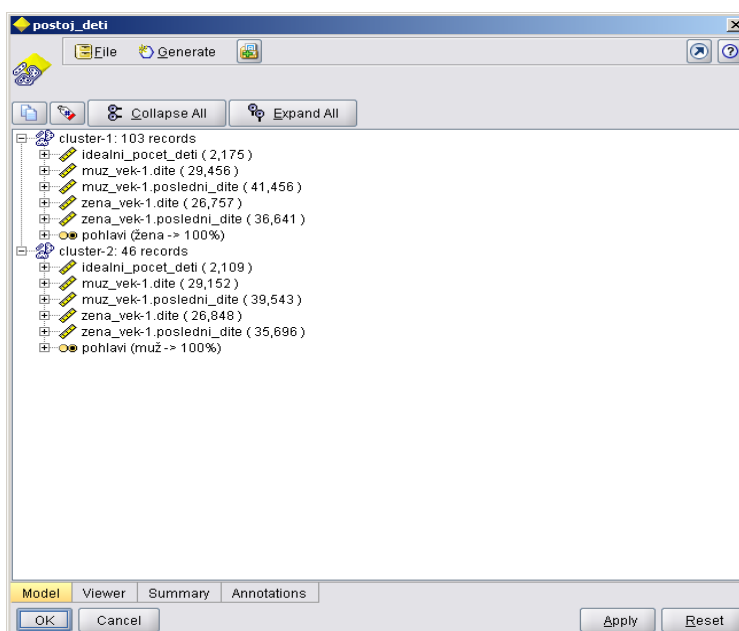
Zdroj: vlastní zpracování



Obrázek 19: Výstup z algoritmu TwoStep pro rok 1997- model 1

Zdroj: vlastní zpracování

Příloha D - Grafické zobrazení výstupů modelů – postoje vysokoškoláků k pořízení dětí



Obrázek 20: Výstup z algoritmu TwoStep pro rok 2011- model 2

Zdroj: vlastní zpracování

The screenshot shows the 'postoj_deti' window with a table of generated rules. The table is sorted by Confidence %.

Consequent	Antecedent	Support %	Confidence %
pohlavi = žena	dite_plnoletost = 5	35,333	64,151
pohlavi = žena	dite_finance = 5	38,0	61,404
pohlavi = žena	dite_ukoncene_vzdelani = 5	63,333	61,053
pohlavi = žena	dite_finance = 5	44,667	59,701
pohlavi = žena	dite_uspory = 3	42,667	59,375
pohlavi = žena	bezdetny_zivot_je_prazdny = 3	38,667	58,621
pohlavi = žena	dite_sarnostatne_bydleni = 4	36,667	58,182
pohlavi = žena	dite_sarnostatne_bydleni = 5	60,667	57,143
pohlavi = žena	dite_dlouha_znamost = 5	42,0	55,556
pohlavi = žena	dite_finance = 5	59,333	55,056
pohlavi = žena	dite_plnoletost = 5	36,667	50,909
pohlavi = žena	dite_dlouha_znamost = 5	47,333	50,704
pohlavi = muž	kdo_chce_deti_mel_by_se_oz...	47,333	49,296
pohlavi = žena	kdo_chce_deti_mel_by_se_oz...	36,667	49,091
pohlavi = muž	dite_plnoletost = 5	59,333	44,944
pohlavi = muž	dite_dlouha_znamost = 5	42,0	44,444
pohlavi = muž	dite_finance = 5	60,667	42,857

Obrázek 21: Vygenerovaná pravidla pro rok 1997-Apriori - model 2

Zdroj: vlastní zpracování

Příloha E - Grafické zobrazení výstupů model – postoje vysokoškoláků k uzavírání manželství

Consequent	Antecedent	Support %	Confidence %
pohlavi = žena	snatek_finance = 3	25,333	68,421
pohlavi = žena	snatek_uspory = 2 zivotni_styl = 3	38,0	66,667
pohlavi = žena	snatek_uspory = 2	40,667	63,934
pohlavi = žena	rozvody = 3	29,333	63,636
pohlavi = žena	snatek_samostatne_b...	30,667	63,043
pohlavi = žena	rozvody = 1 zivotni_styl = 3	36,0	62,963
pohlavi = žena	rozvody = 3 zivotni_styl = 3	28,667	62,791
pohlavi = žena	snatek_samostatne_b... zivotni_styl = 3	30,0	62,222
pohlavi = žena	snatek_dlouha_znam...	49,333	62,162
pohlavi = žena	snatek_dlouha_znam... zivotni_styl = 3	46,667	61,429
pohlavi = žena	snatek_samostatne_b... zivotni_styl = 3	39,333	61,017
pohlavi = žena	rozvody = 1	40,0	60,0
pohlavi = žena	snatek_finance = 4 zivotni_styl = 3	36,667	60,0
pohlavi = žena	zivotni_styl = 3	94,0	58,865
pohlavi = žena	snatek_samostatne_b...	42,0	58,73
pohlavi = žena	snatek_plnoletost = 5 zivotni_styl = 3	38,0	57,895
pohlavi = žena	snatek_finance = 4	39,333	57,627
pohlavi = žena	snatek_dlouha_znam...	31,333	57,447
pohlavi = žena	snatek_dlouha_znam... zivotni_styl = 3	31,333	57,447

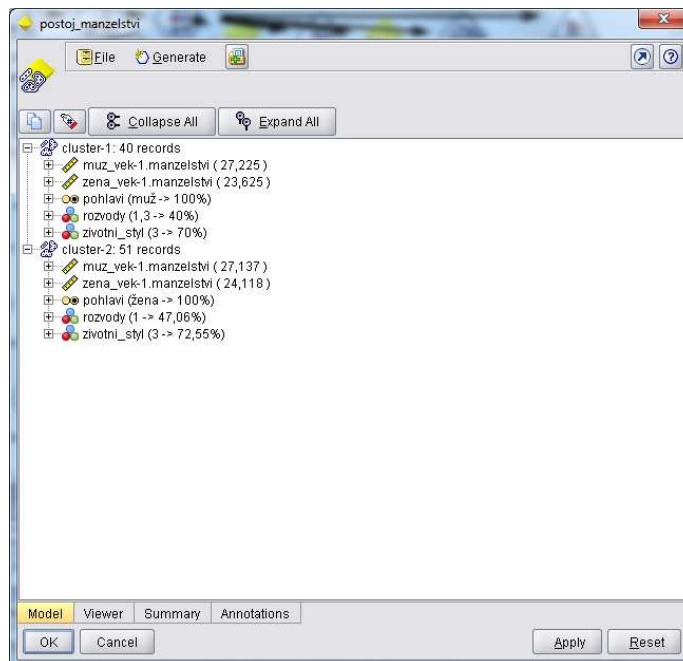
Obrázek 22: Vygenerovaná pravidla pro rok 2011-Apriory – model 3

Zdroj: vlastní zpracování

Consequent	Antecedent	Support %	Confidence %
pohlavi = žena	snatek_finance = 3	25,33	68,42
pohlavi = žena	zivotni_styl = 3 snatek_finance = 3	24,67	67,57
pohlavi = žena	zivotni_styl = 3 snatek_uspory = 2	38,0	66,67
pohlavi = žena	zivotni_styl = 3	94,0	58,87
pohlavi = žena	zivotni_styl = 3 snatek_plnoletost = 4	20,67	51,61
pohlavi = muž	snatek_finance = 5	28,67	51,16
pohlavi = žena	snatek_plnoletost = 4	21,33	50,0
pohlavi = muž	snatek_finance = 4 snatek_samostatne...	21,33	50,0
pohlavi = muž	snatek_plnoletost = 4	21,33	50,0
pohlavi = muž	zivotni_styl = 3 snatek_finance = 5	27,33	48,78

Obrázek 23: Vygenerovaná pravidla pro rok 2011-GRI - model 3

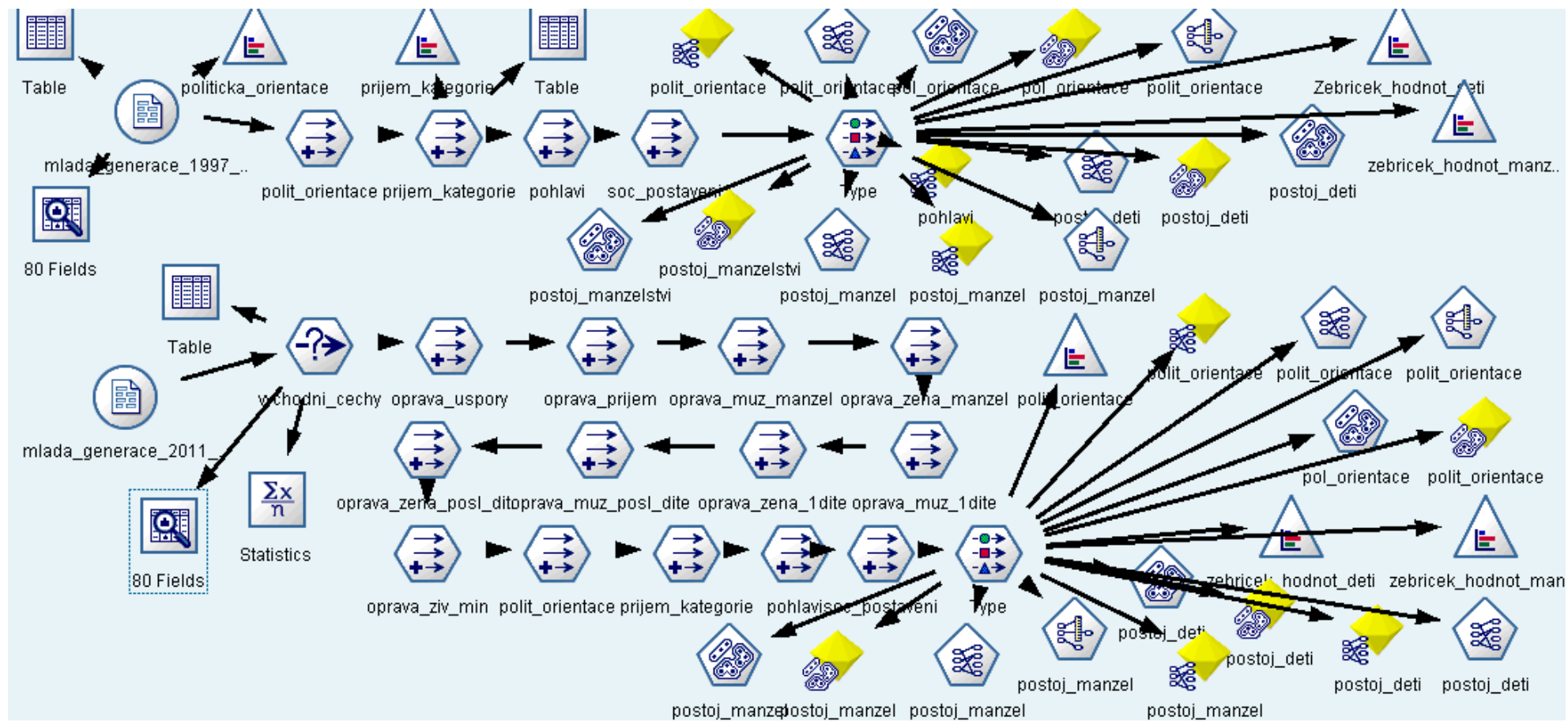
Zdroj: vlastní zpracování



Obrázek 24: Výstup z algoritmu TwoStep pro rok 2011 - model 3

Zdroj: vlastní zpracování

Příloha F - Výsledný stream z programu Clementine



Obrázek 25: Výsledný stream z programu Clementine

Zdroj: vlastní zpracování

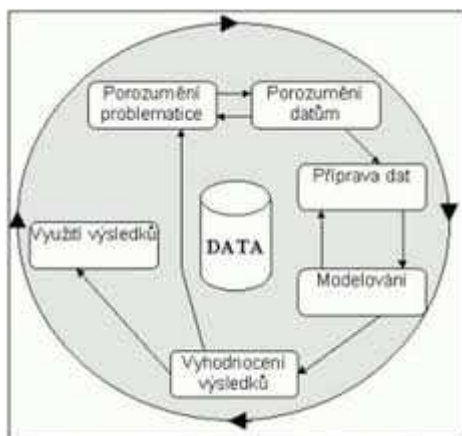
Příloha G - Metodik CRISP a typy dat

Metodika CRISP-DM

Jedná se o metodiku, která vznikla v rámci Evropského výzkumného projektu. Hlavním cílem bylo navrhnout univerzální postup použitelný v různých komerčních aplikacích. Podstatou tedy bylo navrhnout jakýsi standardní model procesu dobývání znalostí v oblasti databází. Vytvořená metodika umožňuje rychlejší, efektivnější, spolehlivější a méně nákladnější řešení rozsáhlých úloh dobývání znalostí. Metodika CRISP-DM kromě jiného také nabízí „průvodce“ možných problémů, které se mohou vyskytnout v komerčních aplikacích, včetně jejich možnosti řešení. Dále nám také umožňuje běžným chybám se vyhnout. [17,32]

Na Evropském výzkumném projektu se podílely firmy NCR (přední dodavatel datových skladů-USA a Dánsko), DaimlerChrysler (Německo), SPSS (tvůrce softwaru Clementine-USA) a OHRA (holandská pojišťovna). Jedná se o firmy, disponující obrovskými zkušenostmi se skutečnými problémy dobývání znalostí z databází. [17,29]

Metodologie CRISP-DM rozčleňuje životní cyklus projektu DM do šesti základních etap, v rámci nichž jsou rozlišovány další kroky. Rozdělení životního cyklu zobrazuje obrázek 26. Pořadí, jak jdou jednotlivé fáze životního cyklu za sebou, není nijak striktně dáno. Volby jednotlivých kroků jsou ovlivňovány výsledky dosažených v předchozích fázích. Velmi často je nutné se k určitým krokům popř. fázím vracet zpět.



Obrázek 26: Metodika CRISP-DM

Zdroj:[17]

Pro popis číselné hodnoty, například řady 0-100 je používán typ data *Range*. Hodnotou může být celé číslo, reální číslo nebo datum (čas). Pro velké množství odlišných hodnot se používá typ *Discrete*. Jedná se o nekonkrétní typ dat, tedy o data u nich použití a informace o hodnotě nejsou známy. Po načtení takovýchto dat bude datový typ příznak, množina nebo bez typu. Pokud proměnná disponuje s dvěma odlišnými hodnotami, například ANO a NE nebo 1 a 2, datový typ této proměnné je *Flag*. Pro popis dat s vícenásobnými odlišnými hodnotami se používá *Set*. Jednotlivé hodnoty jsou zařazeny do této množiny. Jedná se například o mladý/starý. Důležité je to, že nastavení datového typu množina nepřevádí automaticky hodnoty na řetězce. *Order Set* se využívá pro popis dat s vícenásobnými odlišnými hodnotami, které mají vlastní pořadí. Pořadí je definováno přirozeným řazením základních prvků množiny. Jednotlivé číselné hodnoty jsou řazeny podle velikosti. Problematiky řetězových údajů je řešena řazením podle abecedy. V programu se také lze setkat s případem, kdy datový typ bude určen jako *Typeless*. Tento typ se používá pro data, která nelze zařadit do žádného z výše uvedeného typu dat. Toto je využitelné v situaci, zařazení pole do jiného typu znamenalo velké množství různých hodnot položek. [17]

- Porozumění problematice
- Porozumění datům
- Příprava dat
- Modelování
- Vyhodnocení výsledků
- Využití výsledků

Porozumění problematice představuje vstupní fázi, která je zaměřena na definování a pochopení cílů úlohy a požadavků z obchodního hlediska. Požadavky z obchodního hlediska jsou v této fázi převáděny do zadání úlohy pro dobývání znalostí z databází. V této fázi je velmi důležité správně pochopit cíl zadavatele projektu včetně určení nezbytných faktorů, které mohou ve velké míře ovlivnit celkový výstup. A to z toho důvodu, abychom již od počátku nebyli odsouzeni k neúspěchu, z důvodu nepochopení nebo mylného stanovení si cíle. Nezbytná je také definice kritérií pro hodnocení úspěšnosti nebo užitečnosti a efektivnosti získaného výstupu projektu. Jedná se buď o objektivně měřitelná nebo subjektivně vnímatelná kritéria. Před samotným zahájením projektu by měly být známy všechny vstupy, které jsou pro řešení daného problému nezbytné. V rámci fáze porozumění

problematice se mimo jiné také realizuje inventura zdrojů (datových, výpočetních i lidských). Posuzují se především možná rizika, náklady a přínosy a stanovuje se předběžný plán projektu, ve kterém je popsán způsob dosažení cílů data miningu. [32]

Porozumění datům je fáze, která začínám prvotním sběrem dat. Následně se provádí operace za účelem získání základní představy o datech. Jedná se např. o posouzení kvality dat, první „náhled“ do dat, vytipování zajímavých podmnožin záznamů v databázi atd.). Častou činností v této fázi je zjišťování různé deskriptivní charakteristiky dat (četnost hodnot, průměrné hodnoty, minima, maxima,..). Příležitostně se využívají různé vizualizační techniky. [17]

Nejnáročnější fáze je příprava dat, protože data bývají velmi často v odlišných formátech, různých tabulkách, obsahují chybějící hodnoty, odlehlé hodnoty, atd. V rámci této fáze je nezbytné určit, která data budou použita pro následnou analýzu. Na konci této fáze by měla být k dispozici data, která obsahují údaje vyznačené pro danou úlohu a měla by mít podobu, která je požadována následně použitými analytickými algoritmy. Pracnost této fáze dokazuje velmi časté opakování prováděných úkonů a to v nejrůznějším pořadí. Pracnost jednotlivých fází a časová náročnost je zobrazeno níže na grafu 4.

V rámci fáze modelování jsou již aplikovány analytické metody. Důležité je vždy vybrat nejvhodnější, neboť obvykle existuje velký počet různých metod, které lze aplikovat. Důležité je také optimálně nastavit parametry, vzhledem k požadovaným cílům. Nejvhodnější je, na data použít několik různých metod a následně kombinovat jejich výsledky. Použití analytických metod, však velmi často vede k potřebě modifikovat dat, a tedy k návratu k datovým transformacím z fáze Přípravy dat. Velmi často se v této fázi používá metoda shlukování nebo asociační pravidla a to pro hledání „zajímavých“ skupin klientů. Naopak pro rozpoznání rizikových klientů např. z hlediska půjček se využívají algoritmy pro tvorbu rozhodovacích stromů nebo rozhodovacích pravidel. V rámci této fáze se ověřují nalezené znalosti z pohledu metod dobývání znalostí. [17,32]

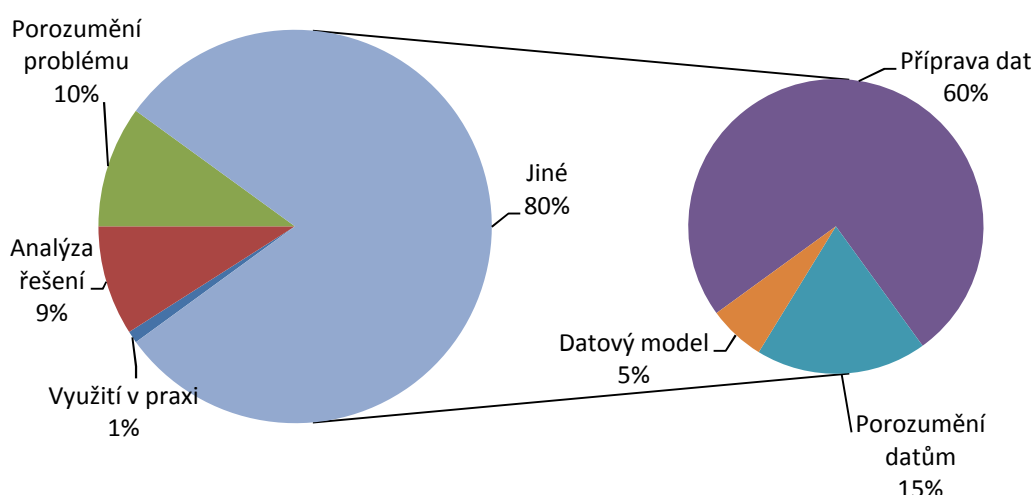
Další fáze životního cyklu projektu DM¹⁰ je hodnocení výsledků. V rámci této fázi se posuzuje úroveň, s jakou model dosáhl stanového cíle a dále se zjišťuje, zda existuje nějaká příčina, díky níž je nutné model považovat za nedostatečný. V případě, že je model označen jako „vyhovující“, následuje precizní kontrola celé data miningové úlohy a stanovuje se, zda

¹⁰ DM- data mining

nebyl přehlédnut nějaký důležitý faktor či úkol. Součástí této kontroly je také ujistění se o kvalitě sestaveného modelu.

Okamžikem, kdy je vytvořen přijatelný model, řešení úlohy ale všeobecně nekončí. Je nutné získané znalosti upravit do podoby přijatelné zadavatelem úlohy-např. manažera. Využití výsledků může představovat jednak pouhé sepsání závěrečné zprávy a na druhé straně implementaci systému pro automatickou klasifikaci nových případů. Zpravidla zákazník a ne analytik provádí kroky vedoucí k využívání výsledků analýzy. [17,32]

Zobrazení časové náročnosti jednotlivých etap DM



Graf 4: Zobrazení časové náročnosti jednotlivých etap DM

Zdroj: upraveno podle [17]

Klasifikace dat

Při klasifikaci dat rozlišujeme dvě třídy:

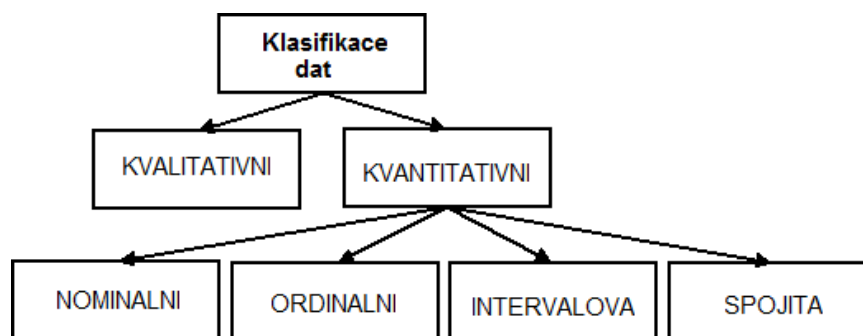
- Kvalitativní
- Kvantitativní

Kvalitativní data se od kvantitativních odlišují tím, že jednotlivé proměnné se odlišují pomocí popisných pojmů. Kvalitativní data se využívají například při segmentaci a

klasifikaci. Příkladem pro kvalitativní proměnou, může být například pohlaví. Pohlaví muž se zpravidla klasifikuje jako „M“, naopak žena jako „Z“. Pro kvantitativní data jsou charakteristické číselné hodnoty. Pokud použijeme znovu příklad pohlaví, tak muž („M“) je možné klasifikovat číselnou hodnotou „1“ a ženu („Z“) číselnou hodnotou „2“. Důležité je však zdůraznit, že tyto číselné hodnoty nemají žádný relativní význam. Kvantitativní data lze využít v rámci prediktivních modelů. Rozlišujeme čtyři typy kvantitativních dat:

- Nominální data
- Ordinální data
- Intervalová data
- Spojitá data

Nominální data reprezentují kategorie a jedná se o číselná data. Výše uvedené číselné hodnoty pro pohlaví (1,2) lze považovat za nominální datové hodnoty. Jak již bylo také výše uvedeno, je důležité, že tyto nominální datové hodnoty nemají relativní význam. Při použití nominální proměnné v modelování by měly být použity pouze dvě hodnoty a to s kódováním 0 a 1. Ordinální data jsou také číselná data, ale od nominálních se liší tím, že jednotlivá data reprezentují kategorie, které mají relativní význam. Tento relativní význam nám tedy umožňuje využít ordinální data pro vyjádření síly nebo důležitosti. Intervalová data jsou reprezentována také číselnými daty, která mají relativní význam a nemají nulový bod. V takovém to případě lze aplikovat různé matematické operace jako sčítání a odčítání. Díky tomu, lze porovnávat hodnocení měřením rozdílů mezi hodnotami. Posledním typem kvantitativních dat jsou data spojitá. Jsou nejčastějším typem dat využívaných při vytváření prediktivních modelů. Tyto data mají nekonečně mnoho hodnot a odpovídají číselnému nepřerušovanému intervalu, číselné ose nebo poloose, všem hodnotám, které v takovém kontinuálním útvaru leží. Je zde smysluplné provádět základní aritmetické operace jako sčítání, odčítání, násobení a dělení. Příkladem takovýchto dat jsou například tržby, zůstatky aj. U sociologické analýzy vystupují spojitá číselná data především jako součtové indexy či průměry položkových hodnot. Grafické zobrazení klasifikace dat zobrazuje obrázek 27. [17, 24]



Obrázek 27: Klasifikace dat

Zdroj: vlastní zpracování

Typy dat v Clementine

V programu Clementine, ve kterém byly aplikovány na data modelovací technik, lze pracovat s několika datovými typy dat. Typy dat, která byla využita v této práci, jsou uvedeny v datovém slovníku. V programu Clementine tedy rozlišujeme:

- Rozsah (Range)
- Diskrétní (Discrete)
- Příznak (Flag)
- Množina (Set)
- Uspořádané množina (Order Set)
- Bez typu (Typeless)

Pro popis číselné hodnoty, například řady 0-100 je používán typ data *Range*. Hodnotou může být celé číslo, reální číslo nebo datum (čas). Pro velké množství odlišných hodnot se používá typ *Discrete*. Jedná se o nekonkrétní typ dat, tedy o data u nich použití a informace o hodnotě nejsou známy. Po načtení takovýchto dat bude datový typ příznak, množina nebo bez typu. Pokud proměnná disponuje s dvěma odlišnými hodnotami, například ANO a NE nebo 1 a 2, datový typ této proměnné je *Flag*. Pro popis dat s vícenásobnými odlišnými hodnotami se používá *Set*. Jednotlivé hodnoty jsou zařazeny do této množiny. Jedná se například o mladý/starý. Důležité je to, že nastavení datového typu množina nepřevádí automaticky hodnoty na řetězce. *Order Set* se využívá pro popis dat s vícenásobnými odlišnými hodnotami, které mají vlastní pořadí. Pořadí je definováno přirozeným řazením základních prvků množiny. Jednotlivé číselné hodnoty jsou řazeny podle velikosti. Problematiky řetězových údajů je řešena řazením podle abecedy. V programu se také lze setkat s případem, kdy datový typ bude určen jako *Typeless*. Tento typ se používá pro data, která nelze zařadit do žádného z výše uvedeného typu dat. Toto je využitelné v situaci, zařazení pole do jiného typu znamenalo velké množství různých hodnot položek. [17]

