

UNIVERZITA PARDUBICE

Fakulta ekonomicko-správní

Analýza přístupů k modelování kvality života

Bc. Jana Svobodová

Diplomová práce

2010

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: 2009/2010

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Jana SVOBODOVÁ**
Studijní program: **N6209 Systémové inženýrství a informatika**
Studijní obor: **Regionální a informační management**
Název tématu: **Analýza přístupů k modelování kvality života**
Zadávací 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 :

Předpokládaným výstupem práce bude analýza přístupů (psychologického, sociologického, ekonomického atd.) ke tvorbě modelu kvality života.

1. Úvod do dané problematiky
2. Výběr faktorů ovlivňující kvalitu života
3. Analýza vybraných přístupů ke tvorbě modelu kvality života
4. Návrh a analýza modelu kvality života

Rozsah grafických prací:

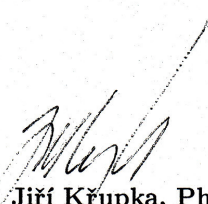
Rozsah pracovní zprávy:

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

Seznam odborné literatury:

- [1] BERKA, P.: Dobývání znalostí z databází. 1. vyd. Praha: Academia, 2003. ISBN 80-200-1062-9
- [2] HAN, J., KAMBER, M.: Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufman Publishers, 2001. 770 s. ISBN 1-55860-901-6
- [3] KAŠPAROVÁ, M., KŘUPKA, J., PÍRKO, J.: Modelování spokojenosti občanů ve vztahu k regionálnímu rozvoji a kvalitě života. Scientific Papers, Series D, Univerzita Pardubice, roč. 13, 2008, s. 109-120. ISBN 978-80-7395-040-8, ISSN 1211-555X
- [4] PAYNE, J.: Kvalita života a zdraví. Alena Malcová. 1. vyd. Praha: TRITON, 2005. 629 s. ISBN 80-7254-657-0

Vedoucí diplomové práce:


doc. Ing. Jiří Křupka, Ph.D.

Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce:

5. října 2009

Termín odevzdání diplomové práce:

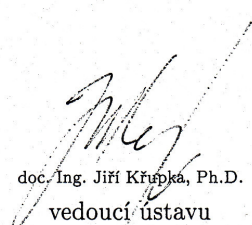
30. dubna 2010



doc. Ing. Renáta Myšková, Ph.D.

děkanka

L.S.


doc. Ing. Jiří Křupka, Ph.D.

vedoucí ústavu

V Pardubicích dne 5. října 2009

Prohlašuji:

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

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

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 23.04.2010

Bc. Jana Svobodová

Poděkování

Ráda bych na tomto místě poděkovala panu doc. Ing. Jiřímu Křupkovi Ph.D. za vedení diplomové práce, cenné rady, odbornou pomoc, připomínky a podporu při jejím vypracovávání. Dále panu Mgr. Janu Mandysovi za poskytnutí cenných informací a konzultací.

Další velký dík patří mým rodičům za jejich celoživotní podporu, pochopení a trpělivost.

SOUHRN

Hlavním cílem této diplomové práce je analýza přístupů k modelování kvality života. Předpokládaným výstupem práce bude analýza přístupů (ekonomického, sociologického, psychologického, apod.) ke tvorbě modelu kvality života. Zvolený model bude navrhnout a analyzován.

KLÍČOVÁ SLOVA

Modelování, Data Mining, rozhodovací stromy, shluková analýza, kvalita života, model kvality života, analýza přístupů, ekonomický přístup, sociologický přístup, psychologický přístup, environmentální přístup, náboženský přístup, zdravotnický přístup, filozofický přístup, návrh modelu, analýza modelu, Maslowova hierarchie potřeb, index lidského rozvoje.

TITLE

Analysis approaches for modelling the quality of life.

ABSTRACT

Principal aim of this thesis is an analysis approaches for modelling the quality of life. The expected output of the work will be to analysis approaches (economic, sociological, psychological, etc.) to create a model quality of life. The chosen model will be designed and analyzed.

KEYWORDS

Modelling, Data Mining, decision trees, cluster analysis, quality of life, quality of life model, analysis approaches, economic approach, sociological approach, psychological approach, environmental approach, religious approach, medical access, philosophical approach, design model, analysis model, Maslow's hierarchy of needs, human development index.

Obsah

ÚVOD	9
1 DEFINICE PROBLÉMU	11
1.1 KVALITA ŽIVOTA	11
1.1.1 Měření kvality života	15
1.1.2 Přístupy ke kvalitě života	19
1.2 MODELOVÁNÍ	22
1.2.1 Model kvality života – zdravotní a environmentální přístup	25
1.2.2 Model kvality života – zdravotní a ekonomický přístup	26
2 VÝBĚR DAT	27
2.1 VSTUPNÍ DATA	27
2.1.1 Zdravotnické ukazatele	27
2.1.2 Environmentální ukazatele	28
2.1.3 Ekonomické ukazatele	29
2.1.4 Doplnující ukazatele	30
2.2 ANALÝZA VSTUPNÍCH DAT	31
2.3 DATOVÝ SLOVNÍK	32
3 PŘÍPRAVA VSTUPNÍCH DAT	33
3.1 POSOUZENÍ KVALITY VSTUPNÍCH DAT	33
3.2 KOREKCE VSTUPNÍCH DAT	35
3.3 PŘEPOČET HODNOT NA 10 000 OBYVATEL	36
3.4 ROZDĚLENÍ VSTUPNÍCH DAT NA TRÉNOVACÍ A TESTOVACÍ MNOŽINU	37
4 NÁVRH MODELU	38
4.1 VLIV ŽIVOTNÍHO PROSTŘEDÍ NA ZDRAVOTNÍ STAV OBYVATEL	38
4.1.1 Rozhodovací stromy C&RT a CHAID	39
4.1.2 Porovnání výsledků vybraných stromů	40
4.2 VLIV ZDRAVOTNÍHO STAVU OBYVATEL NA EKONOMIKU KRAJE	41
4.2.1 Rozhodovací stromy CHAID, QUEST a C&RT	42
4.2.2 Porovnání výsledků vybraných stromů	43
4.3 KVALITA ŽIVOTA ZJIŠTĚNÁ NA ZÁKLADĚ VYBRANÝCH UKAZATELŮ	43
4.4 VÝSLEDNÝ STREAM	45
5 ANALÝZA VÝSLEDKŮ	46
6 ZÁVĚR	49
SYMBOLY A ZKRATKY	51
POUŽITÁ LITERATURA	52

Seznam obrázků

OBRÁZEK 1 - MODEL SYSTÉMU KVALITY ŽIVOTA [ZDROJ: VLASTNÍ]	10
OBRÁZEK 2 - MASLOWOVA HIERARCHIE POTŘEB – ABRAHAM MASLOW -1954 [9]	12
OBRÁZEK 3 - ROZDĚLENÍ HDI V ČLENSKÝCH STÁTECH OSN V ROCE 2009 [8], [34].....	17
OBRÁZEK 4 - RELACE MEZI REÁLNÝM SYSTÉMEM A MODELEM [20]	22
OBRÁZEK 5 - MODEL KVALITY ŽIVOTA - ZDRAVOTNÍ HLEDISKO [ZDROJ: VLASTNÍ].....	25
OBRÁZEK 6 - MODEL KVALITY ŽIVOTA - EKONOMICKÉ HLEDISKO [ZDROJ: VLASTNÍ].....	26
OBRÁZEK 7 - MÍRA REGISTROVANÉ NEZAMĚSTNANOSTI V KRAJÍCH OD ROKU 1998 – 2008 [21]	29
OBRÁZEK 8 - POČET ZJIŠTĚNÝCH TRESTNÝCH ČINŮ V KRAJÍCH OD ROKU 1998 – 2008 [21]	30
OBRÁZEK 9 - ANALÝZA VSTUPNÍCH DAT [ZDROJ: VLASTNÍ]	31
OBRÁZEK 10 - TYPY DAT [ZDROJ: VLASTNÍ]	33
OBRÁZEK 11 - DATOVÉ TYPY ATRIBUTŮ [ZDROJ: VLASTNÍ].....	34
OBRÁZEK 12 - DATA AUDIT [ZDROJ: VLASTNÍ]	34
OBRÁZEK 13 - REKLASIFIKACE ATRIBUTU KRAJ [ZDROJ: VLASTNÍ].....	35
OBRÁZEK 14 - VYŘAZENÍ ATRIBUTŮ [ZDROJ: VLASTNÍ].....	35
OBRÁZEK 15 - PROPOJENÍ DVOU DATOVÝCH SOUBORŮ [ZDROJ: VLASTNÍ].....	36
OBRÁZEK 16 - PŘEPOČET NA 10 000 OBYVATEL [ZDROJ: VLASTNÍ]	37
OBRÁZEK 17 - ROZDĚLENÍ DAT NA TRÉNOVACÍ A TESTOVACÍ MNOŽINU - MATRIX [ZDROJ: VLASTNÍ]	37
OBRÁZEK 18 - ROZDĚLENÍ DO SHLUKŮ - ZDRAVOTNÍ ÚROVEŇ [ZDROJ: VLASTNÍ]	38
OBRÁZEK 19 - VSTUPY A VÝSTUPY ROZHODOVACÍHO STROMU - ZDRAVOTNÍ ÚROVEŇ [ZDROJ: VLASTNÍ].....	39
OBRÁZEK 20- ANALÝZA VÝSLEDKŮ C&RT A CHAID [ZDROJ: VLASTNÍ]	40
OBRÁZEK 21 - ROZDĚLENÍ DO SHLUKU - EKONOMICKÁ ÚROVEŇ [ZDROJ: VLASTNÍ]	41
OBRÁZEK 22 - VSTUPY A VÝSTUPY ROZHODOVACÍHO STROMU [ZDROJ: VLASTNÍ]	42
OBRÁZEK 23 - ANALÝZA VÝSLEDKŮ CHAID, QUEST A C&RT [ZDROJ: VLASTNÍ].....	43
OBRÁZEK 24 - ROZDĚLENÍ DO SHLUKŮ - OBECNÁ KVALITA ŽIVOTA [ZDROJ: VLASTNÍ]	44
OBRÁZEK 25 - VÝSLEDNÝ STREAM [ZDROJ: VLASTNÍ].....	45
OBRÁZEK 26 - ANALÝZA ODHADU ROZHODOVACÍCH STROMŮ [ZDROJ: VLASTNÍ]	46
OBRÁZEK 27 - EVALUAČNÍ GRAF - SROVNÁNÍ ODHADU C&RT A CHAID [ZDROJ: VLASTNÍ]	47
OBRÁZEK 28 - ANALÝZA ODHADU ROZHODOVACÍCH STROMŮ [ZDROJ: VLASTNÍ]	48
OBRÁZEK 29 - EVALUAČNÍ GRAF - SROVNÁNÍ ODHADU CHAID, QUEST A C&RT [ZDROJ: VLASTNÍ].....	48
OBRÁZEK 30 - TYPY DAT [ZDROJ: VLASTNÍ]	55
OBRÁZEK 31 - STATISTICKÁ ANALÝZA VSTUPNÍCH DAT [ZDROJ: VLASTNÍ]	56
OBRÁZEK 32 - KVALITATIVNÍ ANALÝZA VSTUPNÍCH DAT [ZDROJ: VLASTNÍ]	59
OBRÁZEK 33 - ROZDĚLENÍ DO SHLUKŮ - ZDRAVOTNÍ ÚROVEŇ [ZDROJ: VLASTNÍ]	62
OBRÁZEK 34 - DETAIL RS C&RT [ZDROJ: VLASTNÍ]	63
OBRÁZEK 35 - DETAIL RS CHAID [ZDROJ: VLASTNÍ]	63
OBRÁZEK 36 - ROZDĚLENÍ DO SHLUKŮ - EKONOMICKÁ ÚROVEŇ [ZDROJ: VLASTNÍ]	64

OBRÁZEK 37 - DETAIL RS CHAID [ZDROJ: VLASTNÍ]	65
OBRÁZEK 38 - DETAIL RS QUEST [ZDROJ: VLASTNÍ]	65
OBRÁZEK 39 - DETAIL RS C&RT [ZDROJ: VLASTNÍ]	66
OBRÁZEK 40 - ROZDĚLENÍ DO SHLUKŮ - VYBRANÉ UKAZATELE [ZDROJ: VLASTNÍ]	67

Seznam tabulek

TABULKA 1 - HUMAN DEVELOPMENT REPORT 2009 - ŽEBŘÍČEK STÁTŮ [8]	18
TABULKA 2 - STŘEDNÍ DÉLKA ŽIVOTA OD ROKU 1993 – 2008 [21].....	19
TABULKA 3 - DATOVÝ SLOVNÍK – DATOVÝ SOUBOR DATA.CSV [ZDROJ: VLASTNÍ]	32
TABULKA 4 - VSTUPNÍ DATA DATOVÉHO SOUBORU DATA.CSV [3], [4], [21], [29], [30], [32]	54
TABULKA 5 - VSTUPNÍ DATA DATOVÉHO SOUBORU PRŮMĚRY.CSV [ZDROJ: VLASTNÍ]	61
TABULKA 6 - DATOVÝ SLOVNÍK – PRŮMĚRY.CSV [ZDROJ: VLASTNÍ]	61

Seznam rovnic

ROVNICE 1 - VÝPOČET SLOŽEK HDI (VYJMA HDP) [8], [11]	16
ROVNICE 2 - VÝPOČET SLOŽKY HDP PRO HDI [8], [11].....	16
ROVNICE 3 - VÝPOČET ÚROVNĚ VZDĚLANOSTI [8], [11].....	16
ROVNICE 4 - VÝPOČET HDI [8], [11].....	17

ÚVOD

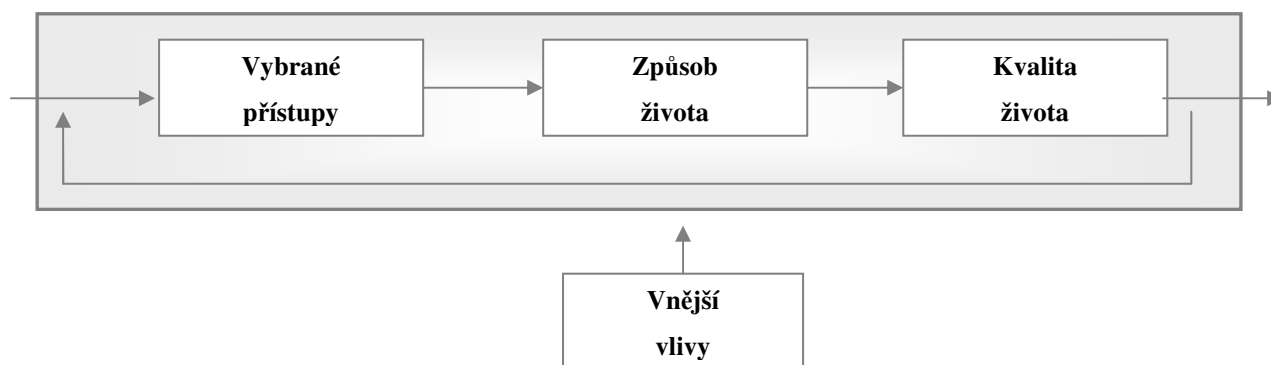
Záměrem mé práce je analýza přístupů k modelování kvality života, vypracování modelu a jeho analýza. V první kapitole se soustředím na vysvětlení jednotlivých základních pojmů, se kterými budu pracovat v této práci. Jelikož je kvalita života velmi složitý a široký pojem, zaměřím se v úvodní části zejména na její definování a na metody, kterými ji můžeme měřit. Dále stručně charakterizuji několik přístupů ovlivňující kvalitu života. Protože se tato práce týká také modelování, závěr této části věnuji vysvětlení principu modelování a jednomu z nejpoužívanějších nástrojů, kterým je Data Mining. Nejvýznamnější část bude věnována analýze přístupů k modelování kvality života pomocí programu Clementine. Zde vypracuji model, který budu následně analyzovat a hodnotit na základě něj kvalitu života v krajích ČR, které jsem si pro tuto práci zvolila.

V České republice bylo od 1. ledna 2000 zřízeno 14 vyšších územních samosprávných celků, jejichž názvy obsahovaly slovo kraj. Krajským orgánem je krajský úřad vykonávající přenesenou působnost státní správy, v jehož čele sedí ředitel. Hlavou každého kraje je hejtman, s výjimkou Prahy, kde je primátor. Praha je také hlavním městem ČR. Tento samosprávný celek disponuje nejmenší rozlohou, člení se na 10 městských obvodů, 22 městských (správních) obvodů a 57 městských částí. Z hlediska průmyslu je v tomto kraji nejvíce zastoupen zpracovatelský a polygrafický průmysl a dále produkce potravin, elektrických a optických přístrojů. Starostou je od roku 2006 Pavel Bém. Praha se nachází přímo uprostřed Středočeského kraje a je zároveň i jeho krajským městem. Středočeský kraj je rozlohou, počtem obcí i obyvatel největším samosprávným celkem v Česku. Hejtmanem kraje je David Rath. Spadá do něj 12 okresů. Stěžejními průmyslovými odvětvími jsou strojírenství, chemie, potravinářství a dále zde sídlí největší česká firma podle tržeb z roku 2008 Škoda Mladá Boleslav. Středočeský kraj sousedí na jihu s Jihočeským krajem, který leží převážně na jihu Čech, ale zasahuje i na Moravu. Krajským městem jsou České Budějovice a hejtmanem je Jiří Zimola. Skládá se ze 7 okresů a převažuje zde zpracovatelský průmysl. Nachází se zde také jaderná elektrárna Temelín. Na severozápadě sousedí s Plzeňským krajem, jehož krajským městem je Plzeň. V čele je hejtmanka Milada Emmerová a je rozdělen na 7 okresů. V kraji převládá strojírenský a potravinářský průmysl. Severozápadně od kraje se nachází další samosprávný celek Karlovarský kraj. Jeho krajským městem jsou Karlovy Vary a hejtmanem Josef Novotný. Kraj je členěn na 3 okresy. Jsou zde významné stáčírny

minerálních vod, sklářské podniky a bohatou tradici zde má také keramický průmysl. Na severovýchod od kraje se nachází kraj Ústecký, jehož krajským městem je Ústí nad Labem a hetmankou je Jana Vaňhová. Spadá do něj 7 okresů. Tento kraj je orientován převážně na těžký průmysl, dále na chemický a papírenský. Na východě leží Liberecký kraj s krajským městem Liberec a hejtmanem Stanislavem Eichlerem. Skládá se ze 4 okresů. Převažuje zde průmysl strojírenský, potravinářský a sklářský. Na severovýchodě Čech se nachází Královéhradecký kraj, jehož krajským městem je Hradec Králové a hejtmanem je zde Lubomír Franc. Je členěn na 5 okresů a převažuje zde zpracovatelský průmysl. Na jih od tohoto celku se nalézá kraj, který zasahuje jak do území Čech, tak i do území Moravy. Jedná se o Pardubický kraj, jehož krajským městem jsou Pardubice a hetmanem je zde Radko Martínek. Rozkládá se do 4 okresů. Největší zastoupení z hlediska průmyslu má průmysl chemický, strojírenský, textilní, oděvní, atd.. Jižním sousedem je pro něj kraj Vysočina, ve kterém je hejtmanem Jiří Běhounek a krajské město je Jihlava. Tento celek se skládá z pěti okresů a převládá zde dřevozpracující, sklářský, strojírenský a potravinářský průmysl. Na jihovýchodě se nachází Jihomoravský kraj, jehož krajským městem je Brno a hejtmanem je Michal Hašek. Člení se na 7 okresů. Je orientován zejména na zpracovatelský průmysl. V severozápadní části Moravy leží Olomoucký kraj, jehož krajským městem je Olomouc a hejtmanem Martin Tesařík. Dělí se na 5 okresů. Dominuje zde zpracovatelsko-strojírenský průmysl. Nejlidnatějším krajem z 14 samosprávných celků je kraj Moravskoslezský. Hejtmanem je zde Jaroslav Palas a krajské město je Ostrava. Kraj se rozkládá do 6 okresů a je orientován zejména na těžký průmysl. Posledním samosprávným celkem je kraj Zlínský, který má v čele hejtmana Stanislava Mišáka. Krajským městem je město Zlín a dělí se na 4 okresy. Zlínský kraj je orientován převážně na zpracovatelský a gumárenský průmysl. [34]

Model systému pro analýzu kvality života jedince či celé společnosti jsem si před samotným začátkem práce navrhla takto:

Obrázek 1 - Model systému kvality života [Zdroj: vlastní]



1 DEFINICE PROBLÉMU

V této kapitole se budu zabývat vymezením základních pojmů, které souvisejí s analýzou přístupů k modelování kvality života. Jedná se zejména o definování pojmu kvality života, zaměřím se na vymezení kvality života na základě Maslowovy hierarchie potřeb, dále uvedu několik pohledů na kvalitu života a v obecné rovině zachyťím různé způsoby jejího měření.

1.1 Kvalita života

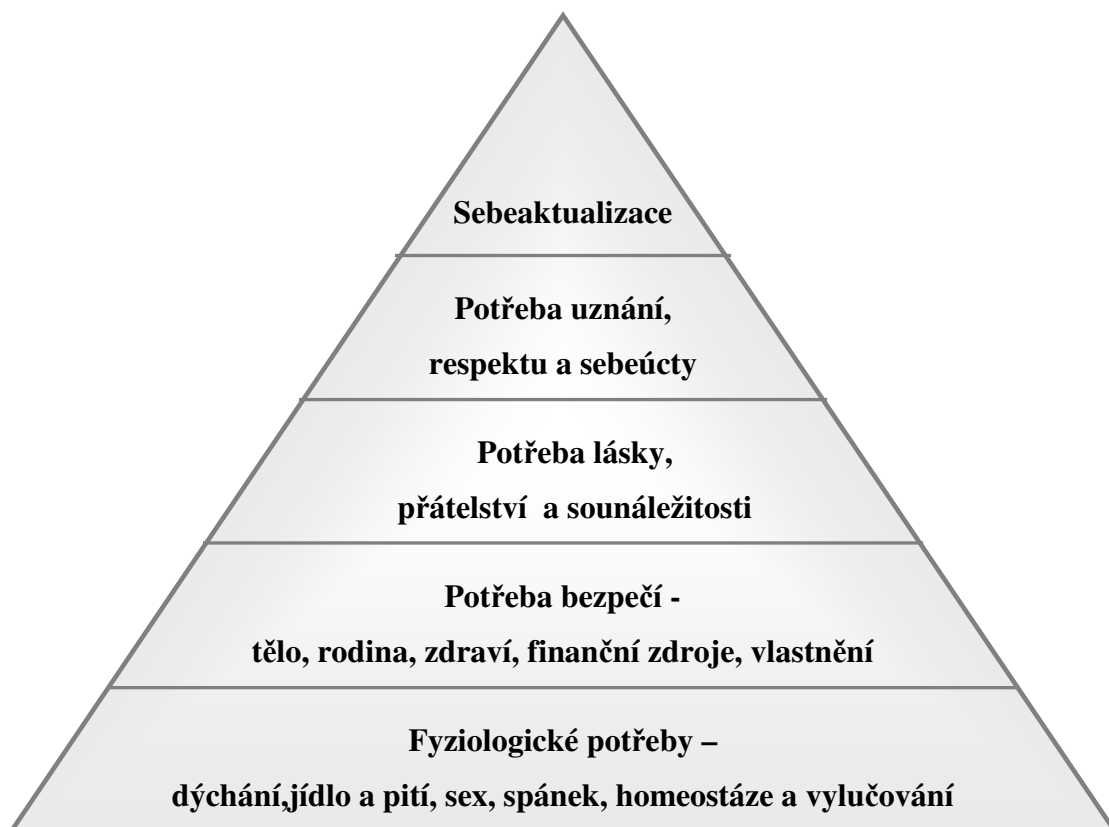
V 80. letech 20. století se začíná pojem kvality života objevovat a zkoumat v různých vědních oborech. Předmětem zájmu byla v té době zejména materiální stránka života společnosti jako celku (založená na objektivizovaném pojetí kvality života), postupně však můžeme sledovat posun k subjektivnímu vnímání a hodnocení kvality života samotného jedince. Smyslem bytí jedince se čím dál víc stává život sám. Představuje přecházení civilizace od extenzivního k intenzivnímu a zejména od kvantitativního (délka života) ke kvalitativnímu (způsob života). V současné době je zkoumání kvality života novým interdisciplinárním oborem, který zjišťuje kvalitu života na různých úrovních od hodnocení životní úrovně národní populace přes porovnání potřeb specifických skupin obyvatel až po měření individuální spokojenosti. [17],[26], [33]

Při studiu kvality života je potřeba brát v úvahu vývoj, změny v prostoru a v čase, společenské souvislosti, historický a kulturní původ, civilizační a generační změny. Významnou roli v chápání kvality života sehrálo a v současné době nadále sehrává, i když v mnohem menší míře než tomu bylo v minulosti, náboženství a víra. Na základě jejich zásad a pravidel jsou lidé často nabádáni k potlačování vlastních potřeb a jejich hodnoty a smysl žití jsou posouvány mimo reálný svět. Silně věřící pokládají náboženství za prvořadé a ostatní hodnoty pro ně nejsou tak významné, dosahování osobních cílů nepovažují za nejdůležitější. [17],[26], [33]

"Hledat smysl existence znamená tedy předně se v tomto chaosu vůbec orientovat a pak vymýšlet a vytvářet cesty, jak se z něho dostat. Kdyby měly trendy pokroku jít i nadále jako doposud, dosáhli bychom posléze toho, že by lidstvo ve svém celku znalo "vše", ale každý jedinec by byl sám pro sebe zase tam, kde byl, než proces dějin civilizace započal." [15]

Definice kvality života vychází z Maslowovy hierarchie potřeb, kterou vidíme na Obrázku č. 2, tj. naplnění základních fyziologických potřeb (potřeba nasycení, spánek, úleva od bolesti) je podmínkou aktualizace a uspokojení potřeb subtilnějších (potřeba bezpečí, potřeba blízkosti jiných, potřeba sebeúcty). Kvalita života jinými slovy představuje uspokojení člověka s dosahováním cílů určující směr jeho života. [6], [9]

Obrázek 2 - Maslowova hierarchie potřeb – Abraham Maslow -1954 [9]



Ve chvíli, kdy máme uspokojeny základní fyziologické potřeby, zaměřujeme se na uspokojení potřeb vyšších. Jestliže na nižší úrovni se nám dostává pocitu nenaplnění, nemůžeme postupovat výš. Jsou-li však naše základní potřeby uspokojeny, přecházíme automaticky na vyšší úroveň a tu předchozí už nevnímáme jako prioritní. Fyziologické potřeby jsou základní potřeby našeho organismu, do kterých patří dýchání, jídlo a pití, sexuální touhy, spánek a vylučování. Rovnovážný stav organismu, ve kterém nám nic nechybí, ale ani nepřebývá, nazýváme homeostáze. V tento okamžik, kdy jsou naše fyziologické potřeby relativně uspokojeny, začneme pohnout po pocitu bezpečí, který si spojujeme s předvídavostí dění ve vnějším světě, snažíme se mít věci pod kontrolou. Toužíme po jistém zaměstnání, rodině, zdraví a finančních jistotách. Česká republika celkem dobře umožňuje uspokojení těchto dvou

základních potřeb a v tom případě nám také umožňuje celkem bezstarostně vstoupit do třetí úrovně Maslowovy hierarchie potřeb. V ní toužíme po citovém naplnění, sounáležitosti, chceme být přijímáni takový jací jsme, máme potřebu milovat a být milováni a není-li tato potřeba naplněna, pocítujeme osamělost, sociální úzkost a depresi. Potřeba sounáležitosti může dokonce překonat naše nižší potřeby (např. „anorektička“ obětuje své fyziologické potřeby a zdraví pro pocit přijetí a sounáležitosti). Na čtvrté úrovni máme potřebu uznání, potřebu respektovat a být respektováni, ale také potřebu sebeúcty. Chceme být přijímáni a pozitivně oceněni, chceme se cítit užiteční, toužíme vynikat ve své profesi a ve svých zájmech. Absence těchto potřeb vyvolává komplexy méněcennosti a nízké sebeúcty. Lidé s nízkou sebeúctou potřebují nepřetržité projevy respektu a uznání všech lidí, přejí si slávu a popularitu. Naopak lidé, kteří se cítí sebevědomí a schopný nevyžadují uznání všech, ale jen od několika málo osob. Psychická nerovnováha může člověku bránit ve vytvoření zdravé sebeúcty. Poslední z Maslowových potřeb souvisejí s růstem naší osobnosti. Ty se od ostatních potřeb liší více méně tím, že jsou potenciálně otevřené do „nekonečna“ – souvisí s dlouhodobou životní motivací. Toužíme se stát tím, kým můžeme nebo máme být. [6], [9]

„Quam bene vivas refert, non quamdiu. Záleží na tom, jak dobře žiješ, nikoliv jak dlouho.“

Seneca

„Život se podobá knize. Blázen v ní letmo listuje, moudrý při čtení přemýšlí, poněvadž ví, že ji může číst jen jednou.“

Johann Paul Friedrich Richter

Na kvalitu života můžeme nahlížet z několika pohledů. Džuka ji rozlišuje na: [5]

- **Objektivní** - představují je životní podmínky jako je příjem, způsob bydlení a množství sociálních kontaktů. Představuje také splnění požadavků týkajících se nemateriálních a sociálních požadavků života, jako výsledku koordinace sociálních, ekonomických zdravotních a enviromentálních podmínek ovlivňujících život lidí.
- **Subjektivní** – představuje ji subjektivní prožívání těchto aspektů (pocit dobrého života a seberealizace, aj.). Týká se všeobecné spokojenosti jedince, ve vztahu k jeho osobním cílům, očekáváním, zájmům, hodnotám a životnímu stylu vůbec.

Podle Křivohlavého lze rozlišovat kvalitu života na základě statického a dynamického pojetí: [13]

- **Statické** – jedná se o stav života lidí k určitému okamžiku (např. kvalita života seniorů v dnešních seniorských zařízeních).
- **Dynamické** – jedná se např. o vývoj kvality života seniorů v posledních patnácti letech.

Tento psycholog také uvádí, že lze rozlišovat různý rozsah kvality života. Tedy, že se lze zabývat kvalitou života jednotlivce, ale také dvojice (manželé, přátelé, žák-učitel, aj.) nebo skupiny lidí a to jak skupiny malé (rodina, třída), tak velké oblasti (obyvatelé Východních Čech, obyvatelé České republiky, aj.).

Poslední dělení, které v této práci uvedu je dělení podle holandských autorů Engela a Bergsma, kteří zkoumali kvalitu života v celé její komplexnosti. Na základě svých poznatků rozlišili tři hierarchicky odlišné sféry: [13], [14], [33]

- **Makro-rovina** – je orientována na otázky velkých společenských celků (dané země, kontinentu, aj.). Život je v tomto pojetí chápán jako absolutní morální hodnota. Kvalita života musí tyto závěry ve své definici plně respektovat.
- **Mezo-rovina** – je orientována na otázky kvality života v tzv. malých sociálních skupinách (např. dětské domovy, domovy důchodců, nemocnice, aj.). V tomto případě nejde pouze o respektování morálních hodnot daného jedince jako tomu bylo v předchozím případě, ale jde zde také o otázky zabývající se sociálním klimatem (např. vztahy mezi lidmi, uspokojování a neuspokojování potřeb, aj.).
- **Personální rovina** - její orientace je soustředěna na život daného jedince (např. život pacienta, život kohokoliv jiného kdo je středem našeho zájmu). Při stanovování kvality života daného jedince, jde o subjektivní hodnocení určité skutečnosti, kdy se může vyjádřit o své individuální kvalitě života, tak jak ji vidí a může zohlednit např. svá očekávání, osobní přání, atd..

1.1.1 Měření kvality života

V této části se budu zabývat některými způsoby měření kvality života. Metod, kterými ji lze měřit je velmi mnoho, a proto zde uvedu pouze obecné rozdělení. Tyto metody měření lze rozdělit do základních tří skupin: [12]

- Hodnotitelem je druhá osoba – ukrývá v sobě mnoho nedostatků a mnohdy se mívá účinkem. Hodnocení velmi často bývá odlišné od hodnocení samotného jedince. To je potřeba mít neustále na paměti. Jednou z metod, kterou zde můžeme použít je Karnofský index (např. lékař vyjadřuje svůj názor na celkový zdravotní stav pacienta k určitému datu).
- Hodnotitelem je sama daná osoba – ke kvalitě života se přistupuje tak, jak je subjektivně definuje ten, kdo je dotazován. Zde se používají např. metody SEIOoL¹ (respondent zaznamenává do dotazníku své životní cíle, hodnotí jejich plnění a určuje jejich důležitost; těchto cílů je právě pět; mezi ně rozdělí 100% svých životních snah, úsilí) a HRQoL (Health Related Quality of Life – zdraví související s kvalitou života).
- Hodnocení vzniklé kombinací předešlých dvou metod – zde lze využít metody MANSA² (hodnotí se nejen celková spokojenost se životem, ale i jeho spokojenost s řadou předem stanovených dimenzí života, které jsou předem přesně stanoveny.) a LSS (Life Satisfaction Scale - škála (stupnice) spokojenosti).

Jedním z důležitých ukazatelů kvality života je v současné době index lidského rozvoje (Human Development Index – HDI), na základě kterého můžeme například srovnávat jednotlivé země z hlediska rozvoje lidského faktoru. Tento index je uveřejňovaný ve známe statistické publikaci Zpráva o lidském rozvoji (Human Development Report) od roku 1990 pod záštitou UNDP (United Nations Development Programme). Tento index je vypočítáván na základě tří kategorií faktorů: [8], [11], [19]

- Lidské zdraví (index očekávané délky života - IODZ)
- Úroveň vzdělanosti (index vzdělání – IV)
- Hmotná životní úroveň (index HDP - IHDP)

¹ SEIOoL – Schedule for the Evaluation of Individual Quality of Life - program hodnocení individuální kvality života

² MANSA – Manchester Short Assessment of Quality of Life - krátký způsob hodnocení kvality života vypracovaný univerzitou v Manchesteru

Pro konstrukci HDI byly u jeho jednotlivých složek stanoveny následující pevné hodnoty minima a maxima: [8], [11]

- průměrná očekávaná délka života při narození: 25 a 85 let;
- gramotnost obyvatelstva staršího 15 let: 0 % a 100 %;
- hrubý podíl populace zapsaných ke studiu 1., 2. a 3. stupně: 0 % a 100 %;
- HDP na hlavu v paritě kup. síly: \$100 a \$40 000.

Jak je vidět z předcházející konstrukce HDI, úroveň vzdělanosti je vyjádřena prostřednictvím dvou ukazatelů a zároveň jsou všechny tyto 4 ukazatele vyjádřeny v různých jednotkách. Proto je potřeba tyto data standardizovat, aby bylo možné data dále porovnávat. Výpočet se provádí na základě níže uvedených vzorců:

Rovnice 1 - Výpočet složek HDI (vyjma HDP) [8], [11]

$$HDI_i = \frac{SH - MIN}{MAX - MIN}$$

Ukazatel HDP na osobu se počítá stejným způsobem až na to, že všechno hodnoty ve zlomku jsou logaritmovány a to z důvodu toho, že se zvyšujícím příjmem stoupají lidské schopnosti (možnosti) pomaleji.

Rovnice 2 - Výpočet složky HDP pro HDI [8], [11]

$$HDI_i = \frac{\log SH - \log MIN}{\log MAX - \log MIN}$$

SH – skutečná hodnota

MIN – minimální hodnota

MAX – maximální hodnota

Jelikož je úroveň vzdělanosti vyjádřena pomocí dvou ukazatelů a pro výpočet HDI je potřeba pouze jeden, musíme ho z těchto ukazatelů vypočítat, to je zachyceno v následující rovnici.

Rovnice 3 - Výpočet úrovně vzdělanosti [8], [11]

Úroveň vzdělanosti = (2/3 x standardizovaná hodnota ukazatele podílu gramotných) + (1/3 x standardizovaná hodnota ukazatele podílu zapsaných ke studiu)

Výpočet konečné hodnoty HDI se provádí na základě následujícího vzorce:

Rovnice 4 - Výpočet HDI [8], [11]

$$HDI = \frac{IODZ + IV + IHDP}{3}$$

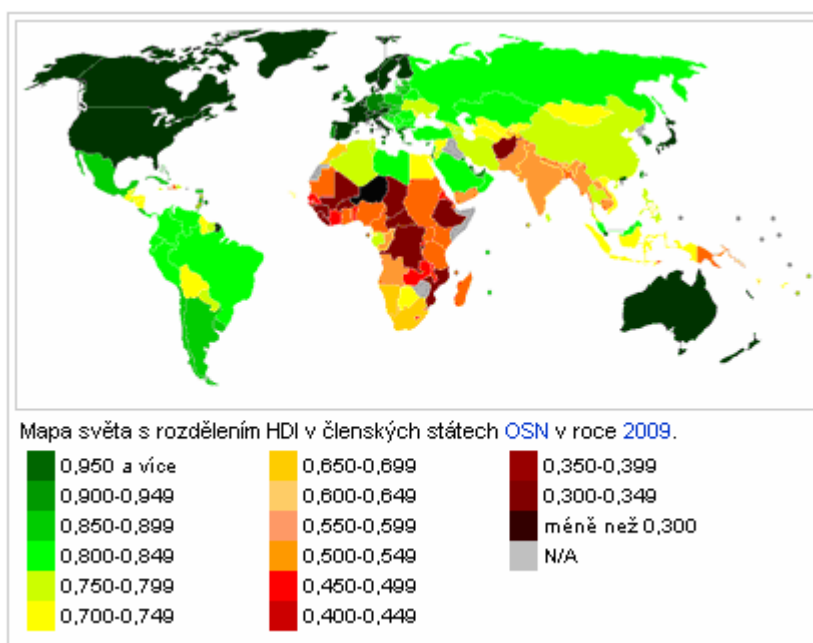
IODZ – index očekávané délky života

IV – index vzdělanosti

IHDP – index HDP

Na následujícím obrázku můžeme vidět rozdělení HDI v jednotlivých členských státech OSN v roce 2009.

Obrázek 3 - Rozdělení HDI v členských státech OSN v roce 2009 [8], [34]



Převážná část Severní Ameriky, Evropy, Skandinávie, Japonska a Austrálie patří mezi země s vyšší úrovní lidského rozvoje. Česká republika se například řadí také do této úrovně. Sice ještě nedosahujeme kvality života jako je například v Rakousku a Německu, ale jsme například před Maďarskem, Polskem a Slovenskem.

Jak můžeme vidět v následující tabulce, v současné době zaujímáme 36. místo.

Tabulka 1 - Human Development Report 2009 - žebříček států [8]

HDI - žebříček států			
1.	Norsko	57.	Seychely
2.	Austrálie	78.	Peru
9.	Švýcarsko	92.	Čína
13.	Spojené státy	130.	Maroko
21.	Velká Británie	143.	Angola
36.	Česká republika	182.	Niger

Jelikož je HDI poměrně dobře korelováno některými ostatními indikátory jako je např. HDP, je velmi zajímavé jejich srovnání. Hodnota HDP v podstatě ukazuje, jak je stát bohatý, zatímco hodnota HDI představuje, do jaké míry je lidsky rozvinut. To tedy znamená, že země, které mají vyšší HDI než HDP, jsou lidsky více rozvinuté, než by odpovídalo jejich bohatství a naopak. Mezi tyto země se například řadí Řecko, Španělsko, Finsko, Kostarika a Thajsko. Naopak mezi země s vyšším HDP a nižším HDI se řadí především ropné státy, dále Švýcarsko a Lucembursko. Co se týče České republiky, tak ta se umístila o něco výše (HDI = 0.87), než by odpovídalo naší hmotné životní situaci.

Na základě indexu lidského rozvoje je možné rozdělit státy do tří skupiny: [11]

- Státy s vysokou úrovní HDI <0.8;1.0>
- Státy se střední úrovní HDI <0.5;0.8)
- Státy s nízkou úrovní HDI (0.5;0.0>

Průvodními ukazateli HDI jsou index lidské chudoby (Human Poverty Index – HPI), který navíc zohledňuje faktory jako je dostupnost vody a dětská podvýživa (HPI-1), populaci pod hranicí chudoby a dlouhodobou nezaměstnanost (HPI-2). Kromě toho se uvádí i index lidského rozvoje zohledňující nerovnost mezi pohlavími (Gender-related Development Index - GDI) a index měřící aktivní účast žen na ekonomickém a politickém životě (Gender Empowerment Measure – GEM), průzkum kvality života v Evropě (European quality of life survey – EQLS), Mercerův celosvětový výzkum kvality života (prováděný společností Mercer Human Resource Consulting) a mnoho dalších. [11]

1.1.2 Přístupy ke kvalitě života

Spokojenost jedince je jedním ze základních předpokladů určující kvalitu jeho života. Ovšem toto uspokojení je ovlivňováno velkou řadou různých přístupů. Tyto přístupy mohou být různého charakteru: [12], [22], [27], [33]

- **Ekonomický přístup** – ekonomická situace člověka či společnosti má čím dál více větší vliv na jeho kvalitu života. Ekonomickou situaci určité společnosti můžeme zachytit mimo jiné také pomocí známého ukazatele HDP. Čím vyšší hodnota tohoto ukazatele, tím lepší ekonomická situace je v dané zemi. Někteří kritici politiky expanzivního hospodářství a neomezeného ekonomického růstu však upozorňovali na negativní dopady na budoucí život člověka.
- **Zdravotní přístup** – jednou z nejvýznamnějších lidských hodnot je zdraví. Zdraví můžeme definovat jako stav úplné tělesné, duševní a sociální pohody, ne pouze jako nepřítomnost nemoci nebo vady. Ovlivňují ho vnitřní (neovlivnitelné) a vnější (ovlivnitelné) vlivy jako jsou např.: genetický základ, vrozené dispozice, věk, pohlaví, životní styl, životní prostředí, lékařská péče, apod.. Z pohledu naděje na dožití určitého věku se věková hranice neustále zvyšuje. To můžeme vidět na následujícím obrázku, kde je zachycen průběh střední délky života od roku 1993 až do roku 2008. Aktuálnější údaje bohužel ještě nejsou k dispozici, ale předpokládá se i nadále zvyšování této věkové hranice.

Tabulka 2 - Střední délka života od roku 1993 – 2008 [21]

STŘEDNÍ DÉLKA ŽIVOTA												
	1993	1995	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
muži	69.2	69.7	71.4	71.6	72.1	72.1	72.0	72.5	72.9	73.4	73.7	74.0
ženy	76.4	76.6	78.1	78.3	78.4	78.5	78.5	79.0	79.1	79.7	79.9	80.1

- **Environmentální přístup** – z hlediska kvality života je tento přístup nejčastěji sledován ve spojení se zdravotním stavem obyvatelstva. Sleduje se zde kvalita ovzduší, pitné vody, hladina hluchosti nebo zatížení organismu cizorodými látkami např. z potravinových řetězců, atd.. V České republice probíhá monitorování zdravotního stavu obyvatelstva ve vztahu k životnímu prostředí prostřednictvím Státního zdravotnického úřadu v Praze (Ústředí Systému monitorování).

- **Psychologický přístup** – s kvalitou života je často spojována osobní pohoda (well-being), kdy se jedná o dlouhodobý emoční stav, ve kterém se odráží spokojenost jedince se životem. Měla by být chápána a měřena prostřednictvím svých kognitivních (životní spokojenost, morálka, aj.) a emočních (pozitivní a negativní emoce) komponent a vyznačuje se konzistencí v různých situacích a stabilitou v čase. Komponentami osobní pohody jsou životní spokojenost (hodnocení toho, jak se jedinci dařilo dosahovat cílů; je orientována do minulosti), morálka ve smyslu mravních zásad (směřuje ke kázni a důvěře; je orientována do budoucna) a jako poslední štěstí (emoční reakce na každodenní život; je orientována do současnosti). Podle Ryffové a Keyesové spojuje v sobě pojetí osobní pohody tyto roviny životní spokojenosti:

- **Sebepřijetí** – pozitivní přijímání své osobnosti;
- **Pozitivní vztahy s druhými** – vřelé uspokojující vztahy s druhými lidmi;
- **Autonomie** – nezávislost na hodnocení a očekávání druhých;
- **Zvládnání životního prostředí** – kompetence pro zvládnání denních nároků;
- **Smysl života** – směřování k dosažení cílů;
- **Osobní rozvoj** – pozitivním směrem vnímaný vývoj osobnosti.

Osobní pohodu může ovlivňovat celá řada faktorů, mezi které například patří zdravotní stav jednice, subjektivní posouzení celkového zdraví, sociálně ekonomický status, věk, zapojení do náboženských aktivit, rodičovství, etnicita, životní události, sociální opora a mnoho dalších.

- **Náboženský přístup** – pro většinu věřících je náboženství hlavním smyslem jejich života. Z hlediska náboženství je pro určité společenské skupiny kvalita života termín rozporuplný, málo akceptovatelný nebo dokonce nepřijatelný. Například předseda papežské Akademie pro život tvrdí, že představa o kvalitě života založené na finančním blahobytu, orientaci na požitky a etickém zesvětštění (sekularismu, omezení církevního vlivu), zatemňuje posvátnost života.

- **Sociologický přístup** – vedle psychologických faktorů na kvalitu života působí také faktory společenské, kterými jsou zejména kultura, vyznání, zdravotní status, příjem, věk, podmínky mobility, uspokojení z práce, atd.. Na kvalitu života zde můžeme nahlížet z několika hledisek. Jedním z prvních hledisek je, že vyjadřuje a operacionalizuje kvalitativní stránky životních procesů a kritéria jejich hodnocení (kvalita života je zde posuzována především podle indikátorů životního prostředí, ukazatelů zdraví a nemoci, úrovně bydlení a rekreace, mezilidských vztahů, volného času, atd.), dále že se jedná o programové politické heslo (obracení pozornosti široké veřejnosti na nové úkoly společnosti, např. změna image USA), sociální hnutí (ekologická hnutí, boje za občanská práva, apod.) a reklamní slogan (objevují se pro spotřebitele nové oblasti konzumu a orientují ho na oblasti prestižní, demonstrativní spotřeby).

- **Filozofický přístup** – zabývá se otázkou dobrého v protikladu ke špatnému. Jinými slovy se zabývá myšlenkou smysluplnosti života. Jedinec si stanoví pro všechno nějaký nadřazený cíl, tyto cíle jsou pro něj potom ukazatelem smysluplnosti jeho života a ukazují mu, co smysl má a co naopak ne. Kvalita života v sobě zahrnuje individuální způsob života, životní podmínky individuální, širších skupin či společnosti jako celku. Zabývá se jí několik filozofických směrů jako jsou například hedonismus³, epikureismus⁴, utilitarismus⁵, kantismus⁶, atd..

³ Hedonismus – pojetí slasti jako nejvyšší hodnoty, dobra

⁴ Epikureismus – ideál duševního klidu a blaženosti, dosažených rozumným prožíváním života

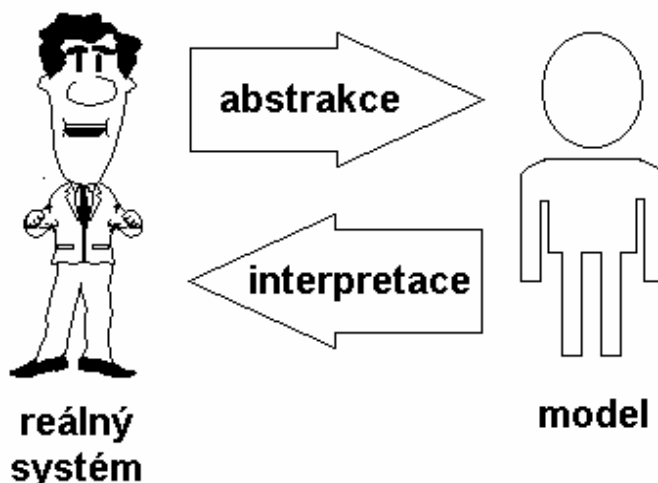
⁵ Utilitarismus – hodnocení jevů z hlediska jejich užitečnosti

⁶ Kantismus – poznáváme jen jevy a ne podstatu věci

1.2 Modelování

Tento pojem reprezentuje jeden z nejobecnějších způsobů zobrazení vnějšího světa, pomocí něhož v něm zkoumáme existující objektivní zákonitosti. Jedná se o experimentální informační proces, při němž se zkoumanému systému (originálu, objektu, dílu,...) přiřazuje podle určitých kritérií jiný systém, fyzický nebo abstraktní, nazývaný model. Model současně představuje strukturu a chování reálného systému. Je propojen dvěma relacemi, kterými jsou abstrakce a interpretace. Pomocí abstrakce uvažujeme pouze nejdůležitější složky reálného systému a ignorujeme méně důležité, interpretace nám objasňuje vztahy mezi modelem a reálným systémem. [20]

Obrázek 4 - Relace mezi reálným systémem a modelem [20]



Modelování ovšem není pouze výsadou vědy a techniky, ale s nejrůznějšími typy modelů se setkávají lidé neustále. Například při řízení auta si řidič vytváří a neustále aktualizuje model celkové dopravní situace zahrnující všechny relevantní faktory. Zvířata si taktéž vytvářejí určitý primitivní model svého okolí, aby mohla účelově reagovat na různé změny situace. [31]

Při modelování musíme dodržovat několik zákonitostí: [31]

- zaměřujeme se na specifické cíle, nesnažíme se vytvořit zcela obecný model;
- volíme optimální úroveň složitosti;
- vyvarujeme se pokušení začlenit do modelu veškeré znalosti o reálném systému;
- nikdy neplánujeme vývoj modelu na dobu delší než 1 rok;
- pokud je to možné, použijeme existující model.

Data Mining (DM) je netriviální získávání implicitních, dříve neznámých a potenciálně užitečných informací z dat. Umožňuje pomocí speciálních algoritmů automaticky objevovat v datech strategické informace. DM je analytická technika pevně spjatá s datovými sklady. Lze ho charakterizovat i jako proces extrakce relevantních, předem neznámých nebo nedefinovaných informací z velmi rozsáhlých databází. K řešení příkladů a tvorbě ukázek byl v této diplomové práci použit programový nástroj firmy SPSS Inc. Clementine. [1], [18]

Rozhodovací stromy (RS) jsou analytické nástroje používané k nalezení pravidel a vztahů v datovém souboru pomocí systematického rozdělování a větvení na nižší úrovně. Jde o učení s učitelem. Jedna z mnoha obecných definic říká, že rozhodovací stromy slouží k postupným hierarchickým rozkladům souboru a jeho vzniklých částí (podsoubory) podle zvoleného optimalizačního kritéria, algoritmu a určených pravidel. Cílem RS je určit takové proměnné, které jsou schopny záznamy rozdělit a snížit tak jejich nejistotu. Problém nastává ve stanovení počtu větví, na které se má dělit každá proměnná. Pokud jsou totiž záznamy podle proměnné rozděleny do příliš mnoha skupin, může se stát, že do každé z těchto skupin náleží pouze několik málo záznamů a nelze tak vyvodit žádná rozhodovací pravidla. [1], [18], [28]

Rozhodovací stromy jsou vhodné pro úlohy, ve kterých má být provedena klasifikace nebo předpověď a jsou dále užitečné v oblastech, ve kterých můžeme hodnoty proměnných rozdělit do relativně malého počtu skupin. Jsou založeny na množství algoritmů: [23], [24], [25]

- CHAID – tento algoritmus dělí skupiny na vhodný počet statisticky homogenních podskupin. Využívá k tomu χ^2 – test závislosti a F – test;
- QUEST – statistický algoritmus, který vybírá proměnné nevychýleně a rychle (vytváří binární stromy);
- C5.0 – vytváří RS na základě algoritmu TDIDT, tak aby v podmnožinách převládaly příklady jedné třídy (tzv. metoda „rozděl a panuj“);
- C&RT) – rozděluje skupinu na dvě části a tím vytváří binární stromy.

Algoritmus TDIDT je obecným algoritmem pro rozhodovací stromy. Nejdříve si na začátku zvolí jeden atribut jako kořen dřívějšího stromu. V druhém kroku data v tomto uzlu rozdělí na podmnožinu podle hodnot zvoleného atributu a přidá uzel pro každou podmnožinu. Ve třetím kroku pokud existuje uzel, pro který nepatří všechny data do téže třídy, pro tento uzel opakuje postup od kroku 1, jinak skončí. Algoritmus funguje pro kategorizovaná data. [1]

Pro výběr nejlepšího větvení se využívá entropie, informační zisk, poměrný informační zisk, Chí kvadrát, Giniho index. [10], [18]

Výhodou RS je jejich schopnost vytvářet srozumitelná pravidla, klasifikaci bez velkých početních požadavků a jsou vhodné pro zpracování jak spojitých tak i kategorických veličin. Naopak jejich nevýhodou je problematičnost pro časové řady a jsou méně vhodné pro odhadnutí úloh, kde je cílem odhad hodnoty nastávající proměnné. [1]

Shluková analýza analyzuje, zda se množina objektů přirozeně rozpadá na výrazné podmnožiny (shluky) objektů si podobných a přitom nepodobných objektům podmnožin ostatních, případně dále analyzuje, jestli existuje celá hierarchie takových rozkladů. Patří mezi metody učení bez učitele. Shlukovací metody můžeme rozdělit podle cíle shlukování na: [16]

- Hierarchické – jedná se o sekvenci vnořených rozkladů, která na jedné straně začíná triviálním rozkladem, kdy každý objekt dané množiny objektů tvoří jednoprvkový shluk, a na druhé straně končí triviálním rozkladem s jedním shlukem obsahující všechny objekty. Na základě směru postupu shlukování dělíme metody na :
 - Aglomerativní – v každém kroku shlukování vybíráme ty dva shluky, které jsou si nejpodobnější a ty pak sloučíme a vytvoříme tak nový shluk;
 - Divizní – za počáteční shluk uvažujeme celou množinu objektů a postupně tak rozdělujeme existující shluky, až jsou všechny jednoprvkové.
- Nehierarchické – na rozdíl od hierarchického shlukování nevytváří hierarchickou strukturu, ale rozkládá výchozí množinu objektů do několika podmnožin takovým způsobem, aby bylo splněno určité kritérium. Cílem je dosažení optimální hodnoty jistého, pro danou metodu, specifického kritéria. Dělíme je na optimalizační metody a metody analýzy módů, které předpokládají určitý předem specifikovaný pravděpodobnostní model. Jedním z těchto metod je algoritmus K-Means.

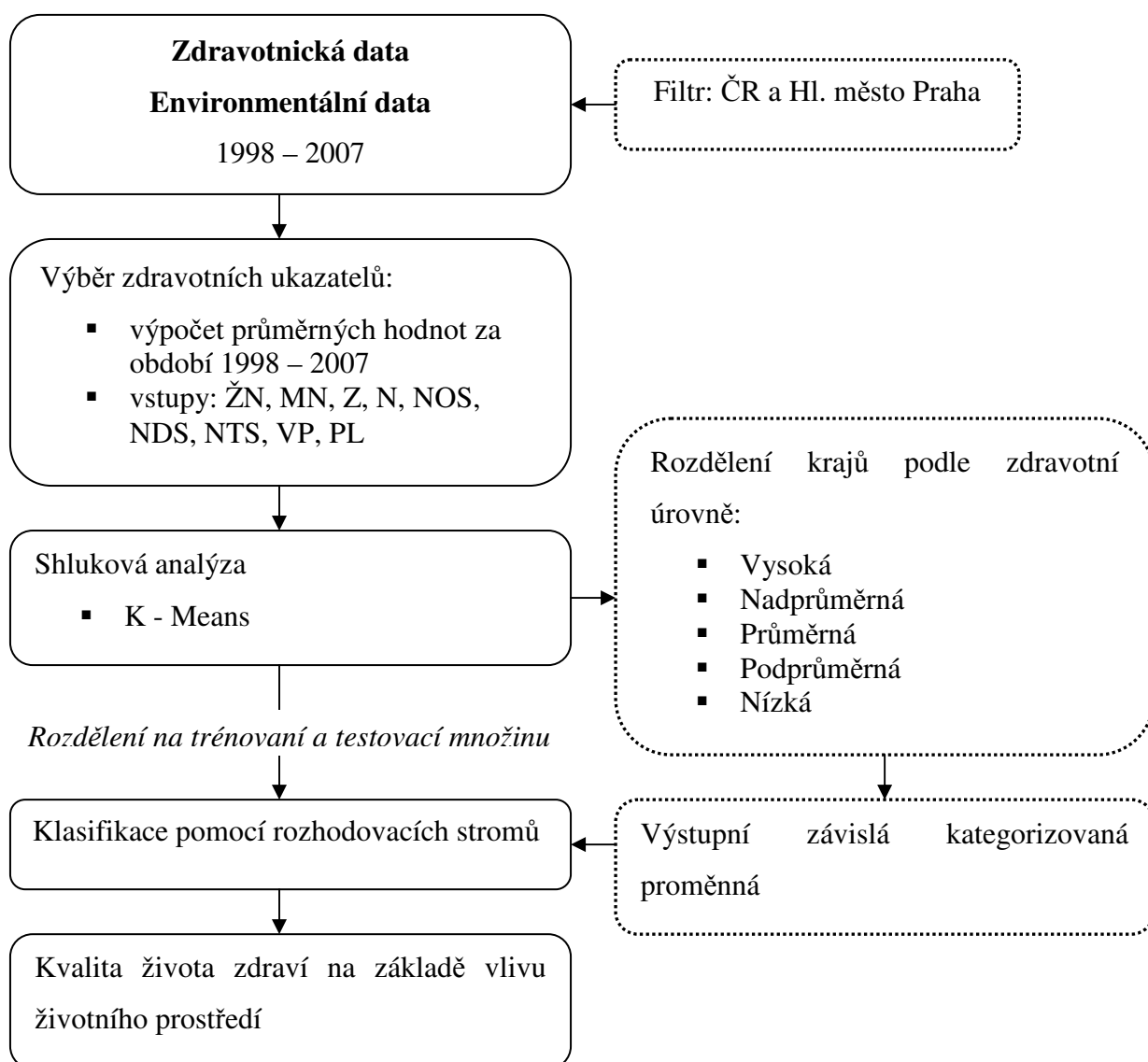
K-means rozdělí množinu dat na k shluků, které následně optimalizuje podle zadaného rozdělovacího kritéria. Algoritmus lze popsat pomocí 4 kroků: [28]

- 1) Rozděl objekty do k neprázdných shluků;
- 2) Spočítej těžiště každého shluku při tomto rozdělení;
- 3) Přiřaď každý objekt do shluku, k jehož těžišti je nejbližší;
- 4) Pokud došlo ke změně přiřazení, opakuj od kroku 2.

1.2.1 Model kvality života – zdravotní a environmentální přístup

Na následujícím obrázku je vidět obecný návrh modelu, pomocí kterého bude zjištěna kvalita života ovlivňovaná životním prostředím. V první řadě nashromáždíme ekonomická a environmentální data od roku 1998 – 2007. Dále ze zdravotnických ukazatelů, po vyfiltrování České republiky a hlavního města Prahy, budou vypočítány průměry hodnot za dané časové období, a ty poslouží jako vstupní nezávislé proměnné do shlukové analýzy. Na základě této analýzy získáme rozdělení krajů podle úrovně a současně závislou výstupní proměnnou. V posledním kroku budou data rozdělena na trénovací a testovací množinu a budou nasazeny rozhodovací stromy, pomocí kterých bude predikována kvalita života.

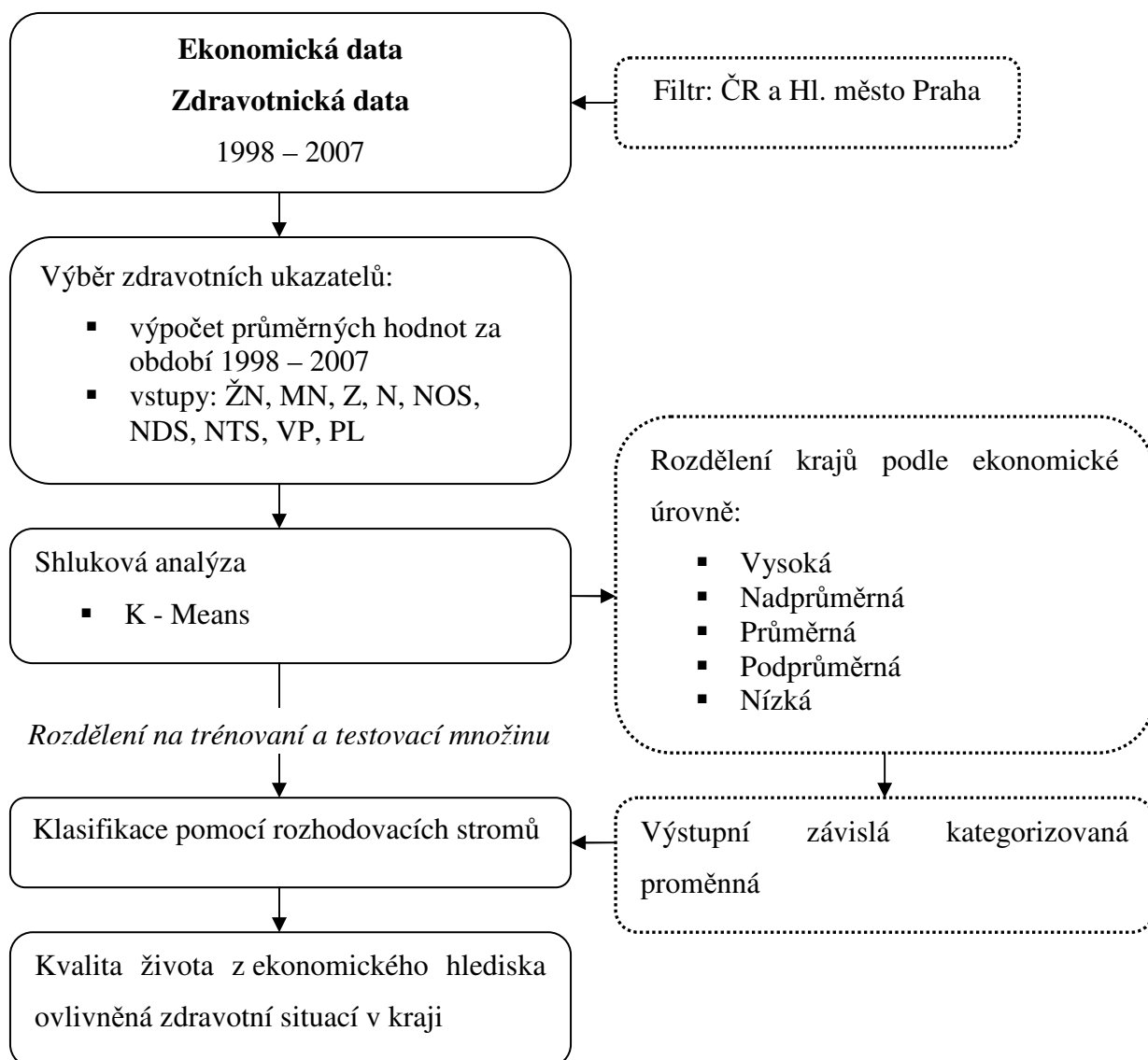
Obrázek 5 - Model kvality života - zdravotní hledisko [Zdroj: vlastní]



1.2.2 Model kvality života – zdravotní a ekonomický přístup

Stejně jako v předchozím případě bude nutné nashromáždit data, v tomto případě pro posouzení kvality života z ekonomického hlediska, kterou ovlivňuje zdravotní stav obyvatel v daném kraji. Sběr dat bude proveden od roku 1998 až 2007 a budou vyfiltrovány data týkající se České republiky a hlavního města Prahy. V dalším kroku budou použita data zprůměrovaných hodnot zdravotnických ukazatelů, jenž zároveň poslouží jako vstupní nezávislé proměnné do shlukové analýzy, kterou rozdělíme kraje podle ekonomické úrovně. Dále budou data rozdělena na trénovací a testovací množinu. V posledním kroku budou nasazeny rozhodovací stromy, na základě kterých bude odhadnuta kvalita života v jednotlivých krajích.

Obrázek 6 - Model kvality života - ekonomické hledisko [Zdroj: vlastní]



2 VÝBĚR DAT

Tato kapitola je věnována sběru vstupních dat a jejich popisu. Jak již bylo řečeno, kvalita života bude sledována za kraje České republiky. Jelikož se tato práce zabývá analýzou přístupů k modelování kvality života, byly zvoleny přístupy, na základě kterých bude analýza probíhat. Jedná se o analýzu kvality života z hlediska zdraví jedince na základě životního prostředí, ve kterém žije a v druhém případě vycházíme z předpokladu, že pokud jsou jedinci v daném kraji zdravější, mohou i více pracovat a tím mohou vytvářet příznivější ekonomické podmínky v daném kraji. Výběr vstupních dat je zachycen v Příloze č. 1.

2.1 Vstupní data

Vstupní data byla získána za pomoci webového portálu RISY.CZ (Portál regionální informačních servisů), Českého statistického úřadu Pardubického kraje, Veřejné databáze ČSÚ ČR, ÚZIS – Ústavu zdravotnických informací a statistiky ČR a Českého hydrometeorologického ústavu a to za období 1998 – 2007. Obsahují 29 atributů z oblasti zdravotnictví, životního prostředí, ekonomiky a další doplňující ukazatele.

2.1.1 Zdravotnické ukazatele

Z oblasti zdravotnictví jsem vybrala tyto základní ukazatele:

- **Naděje na dožití** (muži a ženy) – tento ukazatel vyjadřuje počet roků, který v průměru prožije osoba právě x-letá za předpokladu, že po celou dobu jejího dalšího života se nezmění řád vymírání, zjištěný úmrtnostní tabulkou, zkonstruovaný pro daný kalendářní rok nebo jiné období. Tomuto ukazateli se také někdy říká Střední délka života. Zjišťuje se pro muže a ženy zvlášť a pro menší územní celky se většinou vypočítává za pětileté období z důvodu vyloučení nahodilých výkyvů; [2]
- **Zemřelí na novotvary** – nebo-li nádorové onemocnění. Společným rysem novotvarů je určitá populace buněk, která se vymkne kontrole a začne velmi rychle autonomně růst. Může se jednat např. o zhoubné novotvary prsu, děložního těla, hrtanu, žlučníku, tlustého střeva, žaludku, atd.. V tomto případě tento ukazatel vyjadřuje počet jedinců, kteří zemřeli na novotvar;
- **Živě narození** – tento ukazatel je také nazýván obecnou mírou plodnosti, vyjadřuje počet hlášení o živě narozených dětí;
- **Mrtvě narození** – ukazatel vyjadřuje počet hlášení mrtvě narozených dětí;

- **Zemřelí** – ukazatel vyjadřuje celkový počet zemřelých v určitém roce;
- **Zemřelí na nemoci oběhové soustavy** – vyjadřuje počet osob, kteří zemřeli na nemoci oběhové soustavy. Jsou to například osoby zemřelí na cévní nemoci mozku, plicní embolii, angínu pectoris, akutní infarkt myokardu, atd.;
- **Zemřelí na nemoci dýchací soustavy** – jedná se ukazatel znázorňující počet osob, kteří zemřeli na některou z nemocí dýchací soustavy jako je např.: zánět plic, rozedma plic, akutní infekce horních cest dýchacích, apod.;
- **Zemřelí na nemoci trávicí soustavy** – vyjadřuje počet osob, kteří zemřeli na nemoci trávicí soustavy. K těmto nemocím patří nemoci appendixu, žaludeční vředy, vředy dvanáctníku, atd.;
- **Zemřelí na vnější příčiny** – vyjadřuje počet osob, kteří zemřeli na vnější příčiny. Řadí se sem například sebevraždy, atd..
- **Počet nemocnic** – jaký je stav nemocničních zařízení v daném kraji v průběhu let 1998 – 2007;
- **Počet lékařů** - jaký je stav lékařů v daném kraji v průběhu let 1998 – 2007.

2.1.2 Environmentální ukazatele

Ukazatele životního prostředí byly jako základní zvoleny Pořízené investice na ochranu životního prostředí a Emise základních znečišťujících látek REZZO⁷ 1 – 3 (stacionární zdroje):

- **Pořízené investice na ochranu životního prostředí** – jedná se o investice na ochranu životního prostředí kraje, bez ohledu na to, jakou částí se na jeho realizaci kraj podílí;
- **Emise základních znečišťujících látek:**
 - tuhé látky - prachové částice;
 - oxid siřičitý (SO²) – zdrojem jsou např. elektrárny;
 - oxidy dusíku (NO_x) – zdrojem jsou např. automobily a lokální topeniště;
 - oxid uhelnatý (CO) - jehož hlavním zdrojem je silniční doprava a spalování fosilních paliv.

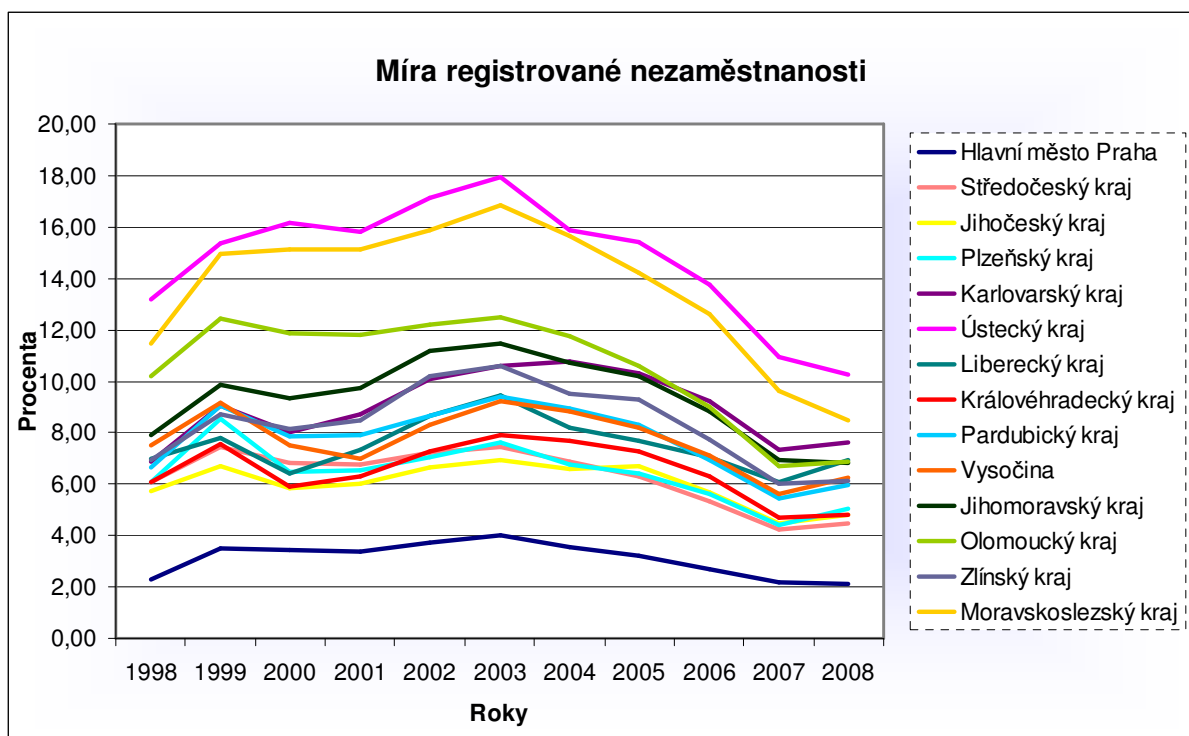
⁷ REZZO 1 – Zvláště velké a velké zdroje znečišťování ovzduší, REZZO 2 – Střední zdroje znečišťování, REZZO 3 – Malé zdroje znečišťování [29]

2.1.3 Ekonomické ukazatele

Ke zjištění ekonomické situace kraje byly vybrány tyto ukazatele:

- **Hrubý domácí produkt** – tento ukazatel se používá pro souhrnné hodnocení zemí (regionů) z hlediska jejich ekonomické výkonnosti;
- **Hrubá mzda** – jedná se o průměrnou hrubou mzdu poskytovanou zaměstnavatelem zaměstnanci v daném kraji. Skládá se ze základní mzdy + osobního ohodnocení, příplatků, prémie a odměn + náhrad mezd. Zahrnuje v sobě dále zákonné srážky (záloha na daň z příjmu, zdravotní a sociální pojištění);
- **Čistý disponibilní důchod domácností** - je to částka, kterou mohou domácnosti věnovat na konečnou spotřebu, na úspory finančních aktiv a na kumulaci hmotných i nehmotných aktiv. Tento ukazatel nám jinými slovy naznačuje úroveň materiálního bohatství domácností trvale bydlících v jednotlivých regionech;
- **Míra registrované nezaměstnanosti** – tento ukazatel vyjadřuje podíl počtu nezaměstnaných registrovaných úřady práce na disponibilní pracovní síle (v %). Průběh tohoto ukazatele za období 1998 – 2008 je zachycen na následujícím obrázku.

Obrázek 7 - Míra registrované nezaměstnanosti v krajích od roku 1998 – 2008 [21]

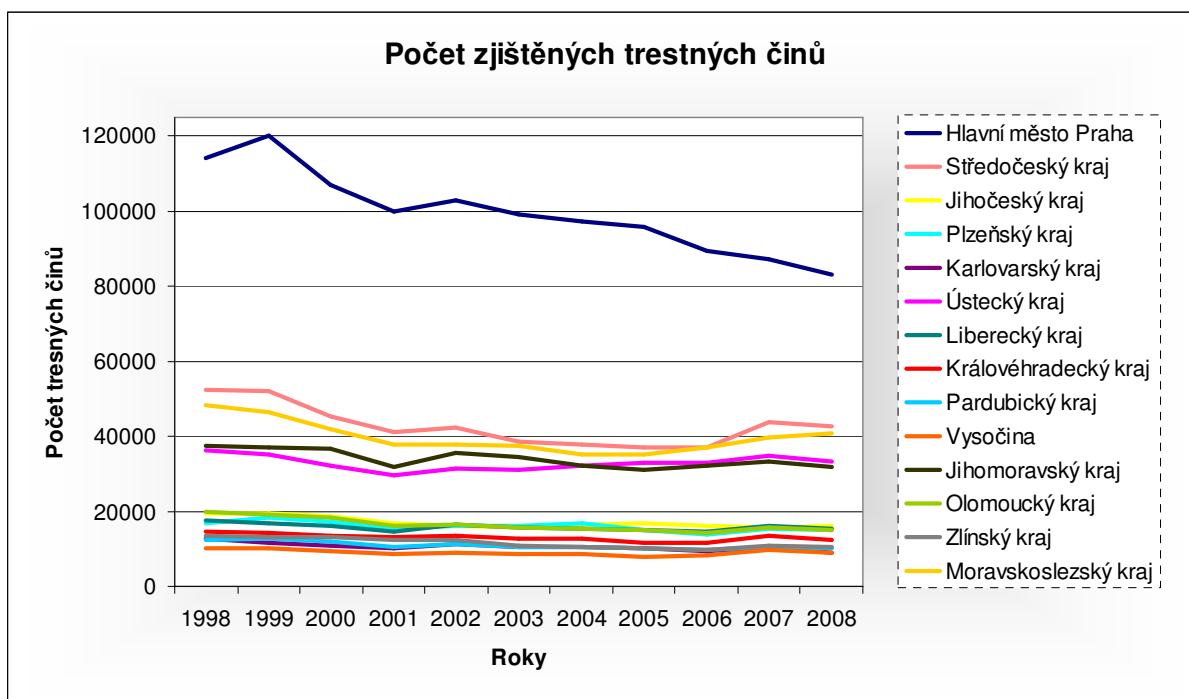


2.1.4 Doplnující ukazatele

Jedná se o ukazatele přímo charakterizující dané oblasti:

- **Kraj** – tento ukazatel znázorňuje číselné označení samosprávných celku;
- **Stav obyvatelstva** – jedná se o aktuální stav obyvatelstva v jednotlivých letech počítaný k 31.12. daného roku;
- **Rozloha** – představuje rozlohu kraje v daném roce;
- **Celkový přírůstek obyvatelstva** – jedná se o součet přirozeného přírůstku obyvatel a migračního salda za dané období;
- **Sňatky** – udává počet hlášení o uzavření manželství;
- **Rozvody** – udává počet hlášení o rozvodu manželství;
- **Kriminalita** – vyjadřuje počet zjištěných trestných činů v daném kraji. Trend lze vysledovat z níže uvedeného obrázku.

Obrázek 8 - Počet zjištěných trestných činů v krajích od roku 1998 – 2008 [21]



Všechny ukazatele jsou sledovány ročně a to v období 1998 – 2007 (platí pro datový soubor, ne pro předchozí grafické přílohy).

2.2 Analýza vstupních dat

Všechna data se nachází v jedné tabulce souboru typu xls. Pro potřeby programu Clementine byl tento soubor převeden do formátu csv, poté byla data importována pomocí uzlu „Var. File“. Samotná analýza dat byla provedena pomocí uzlu „Statistics“, který zjistí u každé proměnné počet hodnot (Count), průměr (Mean), minimum (Min), maximum (Max), medián (Median) a modus (Mode). Modus, neboli hodnota, která se v daném souboru vyskytuje nejčastěji, je v tomto datovém souboru u některých ukazatelů několikrát, proto je zde vypisována hodnota nejmenší, která se nejčastěji opakuje. Pokud se zde hodnoty vyskytují pouze jednou, je vypisována vždy ta nejmenší. Celkovou analýzu je možné vidět v Příloze č. 3. Ukázkou můžete vidět na Obrázku č. 9:

Obrázek 9 - Analýza vstupních dat [Zdroj: vlastní]

---Stav obyvatelstva

Statistics	
Count	150
Mean	1367938.667
Sum	205190800.000
Min	303714
Max	10381130
Range	10077416
Median	625993
Mode	303714*

*Multiple modes exist. The smallest value is shown.

---Rozloha

Statistics	
Count	150
Mean	10515.440
Sum	1577316.000
Min	495.800
Max	78868.400
Range	78372.600
Median	5334.900
Mode	4518.500

---Celkový přírůstek obyv.

Statistics	
Count	150
Mean	1553.653
Sum	233048.000
Min	-25591
Max	93941
Range	119532
Median	-148.500
Mode	-930*

*Multiple modes exist. The smallest value is shown.

2.3 Datový slovník

Datový slovník byl vytvořen v programu Excel a má následující podobu:

Tabulka 3 - Datový slovník – datový soubor DATA.CSV [Zdroj: vlastní]

č.	Ukazatel	Typ proměnné	Type	Rozsah	Měrná jedn.	Definice ukazatele
1	Rok	kategorizovaná	Ordered set	<1998;2007>	rok	označení období
2	Kraj	kategorizovaná	set	<1;15>	označení	označení oblasti (1-ČR;2-Hl.město Praha;3-Středočeský kraj;4-Jihočeský kraj;5-Plzeňský kraj;6-Karlovarský kraj;7-Ústecký kraj;8-Liberecký kraj;9-Královéhradecký kraj;10-Pardubický kraj;11-Vysočina;12-Jihomoravský kraj;13-Olomoucký kraj;14-Zlínský kraj;15-Moravskoslezský kraj)
3	Stav obyvatelstva	spojitá	range	<303714;10381130>	osoby	aktuální stav obyvatelstva k 31.12.
4	Rozloha	spojitá	range	<495.8;78868.4>	km ²	rozloha kraje
5	Celkový přírůstek obyv.	spojitá	range	<-25591;93941>	osoby	přirozený přírůstek obyvatel + migrace
6	Sňatky	spojitá	range	<1573;57157>	osoby	počet hlášení o uzavření manželství
7	Rozvody	spojitá	range	<898;33060>	osoby	počet hlášení o rozvodu manželství
8	Kriminalita celkem	spojitá	range	<7993;426626>	trest. činy	počet zjištěných trestných činů
9	Naděje na dožití - M	spojitá	range	<68.8;75.6>	věk	střední délka života mužů
10	Naděje na dožití - F	spojitá	range	<76.1;81.1>	věk	střední délka života žen
11	Hrubá mzda	spojitá	range	<10214;27805>	Kč	průměrná hrubá mzda
12	ČDDD	spojitá	range	<94838;228853>	Kč/obyv.	čistý disponibilní důchod domácností
13	HDP	spojitá	range	<154569;736228>	na 1 obyv.	hrubý domácí produkt
14	MRN	spojitá	range	<2.160;17.940>	%	počet nezaměstnaných registrovaných úřady
15	Živě narození	spojitá	range	<2777;114632>	osoby	počet hlášení o živě narozených
16	Mrtvě narození	spojitá	range	<6;315>	osoby	počet hlášení o mrtvě narozených
17	Zemřelí	spojitá	range	<2851;111288>	osoby	celkový počet zemřelých
18	Novotvary	spojitá	range	<819;29364>	osoby	počet zemřelých na novotvary
19	Nemoci OS	spojitá	range	<1240;60397>	osoby	počet zemřelých na nemoci oběhové s.
20	Nemoci DS	spojitá	range	<104;6040>	osoby	počet zemřelých na nemoci dýchací s.
21	Nemoci TS	spojitá	range	<115;4823>	osoby	počet zemřelých na nemoci trávicí s.
22	Vnější příčiny	spojitá	range	<169;7295>	osoby	počet zemřelých na vnější příčiny
23	Počet nemocnic	spojitá	range	<5;225>	zařízení	aktuální počet nemocnic
24	Počet lékařů	spojitá	range	<1021;43676>	osoby	aktuální počet lékařů
25	PIO ŽP	spojitá	range	<227951;44435229>	v tis. Kč	investice na ochranu životního prostředí
26	Tuhé látky	spojitá	range	<535.7;75699.6>	t/rok	emise tuhých látek do ovzduší
27	SO ₂	spojitá	range	<1422.3;432038.2>	t/rok	emise oxidu siřičitého do ovzduší
28	NO _x	spojitá	range	<1751.6;164223.1>	t/rok	emise oxidů dusíku do ovzduší
29	CO	spojitá	range	<2117.8;397930.2>	t/rok	emise oxidu uhelnatého do ovzduší

3 PŘÍPRAVA VSTUPNÍCH DAT

Tato kapitola je věnována přípravě dat, jejich korekci, přetypování, přepočítání, reklasifikaci a rozdělení atributů na trénovací a testovací množinu. Příprava dat bude provedena pomocí programu Clementine a MS Excel. Celkové výstupy lze zjistit v Příloze č. 3, 4, 5 a 6.

3.1 Posouzení kvality vstupních dat

Pro následující práci je nutné posouzení kvality vstupních dat. V tomto případě byly nadefinovány atributy Rok a Kraj jako Range. Na následujícím obrázku je vidět jejich přetypování na Set, tedy množinu u atributu Kraj a Order Set u atributu Rok, neboli přetypování na uspořádanou množinu. Všechny ostatní atributy byly nadefinovány správně.

Obrázek 10 - Typy dat [Zdroj: vlastní]

Field	Type	Values	Missing	Check	Direction
Rok	Ordered Set	1998,199...		None	In
Kraj	Set	"1","10","1...		None	In
Stav obyvatel...	Range	[303714,1...		None	In
Rozloha	Range	[495.8,78...		None	In
Celkový přírů...	Range	[-25591,9...		None	In
Sňatky	Range	[1573,571...		None	In
Rozvody	Range	[898,33060]		None	In
Kriminalita c...	Range	[7993,426...		None	In
Naděje na d...	Range	[68.8,75.6]		None	In
Naděje na d...	Range	[76.1,81.1]		None	In
Hrubá mzda	Range	[10214.0,...		None	In
ČDDD	Range	[94838,22...		None	In
HDP	Range	[154569,7...		None	In
MRN	Range	[2.16,17.94]		None	In
Živě narození	Range	[2777,114...		None	In
Mrtvě narození	Range	[6,315]		None	In
Zemřelí	Range	[2851,111...		None	In

V dalším kroku bylo potřeba zkontrolovat, zda program správně nadefinoval datové typy atributům. Jak je vidět z následujícího obrázku, bylo nutné přetypovat jeden atribut, který byl nadefinován jako proměnná, se kterou by se dále pracovalo jako s číslem, Tedy datový typ Integer bylo potřeba změnit na datový typ String. Na Obrázku č. 11 lze tuto změnu vysledovat zaškrtnutým políčkem Override. V tomto případě ale jednotlivá čísla představují kategorie, do kterých jsme si kraje rozdělily, nepředstavují číselné hodnoty.

Obrázek 11 - Datové typy atributů [Zdroj: vlastní]

Field	Override	Storage	Input Format
Rok	<input type="checkbox"/>	Integer	
Kraj	<input checked="" type="checkbox"/>	String	
Stav obyvatelstva	<input type="checkbox"/>	Integer	
Rozloha	<input type="checkbox"/>	Real	
Celkový přírůstek obyv.	<input type="checkbox"/>	Integer	
Sňatky	<input type="checkbox"/>	Integer	
Rozvody	<input type="checkbox"/>	Integer	
Kriminalita celkem	<input type="checkbox"/>	Integer	
Naděje na dožití - M	<input type="checkbox"/>	Real	
Naděje na dožití - F	<input type="checkbox"/>	Real	
Hrubá mzda	<input type="checkbox"/>	Real	
ČDDD	<input type="checkbox"/>	Integer	
HDP	<input type="checkbox"/>	Integer	
MRN	<input type="checkbox"/>	Real	
Živě narození	<input type="checkbox"/>	Integer	
Mrtvě narození	<input type="checkbox"/>	Integer	
Zemřelí	<input type="checkbox"/>	Integer	
Nonotvary	<input type="checkbox"/>	Integer	
Nemoci OS	<input type="checkbox"/>	Integer	
Nemoci DS	<input type="checkbox"/>	Integer	

Pro vstupní analýzu dat je možné kromě uzlu „Statistics“ využít také uzel „Data audit“, pomocí kterého lze zjistit, zda jsou data validní či ne. Tedy, jestli datový soubor obsahuje skutečně všechna data. Jak je vidět z následujícího obrázku (popř. přílohy č. 4), datový soubor je validní, jelikož každý sloupec obsahuje 150 datových záznamů.

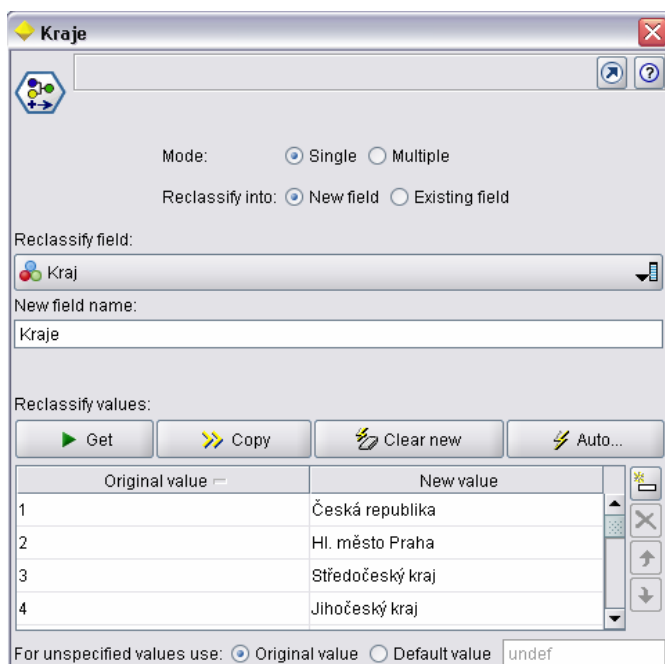
Obrázek 12 - Data audit [Zdroj: vlastní]

Field	Graph	Type	Min	Max	Valid
Rok		Ordered Set	1998	2007	150
Kraj		Set	--	--	150
Stav obyvatelstva		Range	303714	10381130	150
Rozloha		Range	495.800	78868.400	150
Celkový přírůstek obyv.		Range	-25591	93941	150
Sňatky		Range	1573	57157	150
Rozvody		Range	898	33060	150
Kriminalita celkem		Range	7993	426626	150
Naděje na dožití - M		Range	68.800	75.600	150
Naděje na dožití - F		Range	76.100	81.100	150

3.2 Korekce vstupních dat

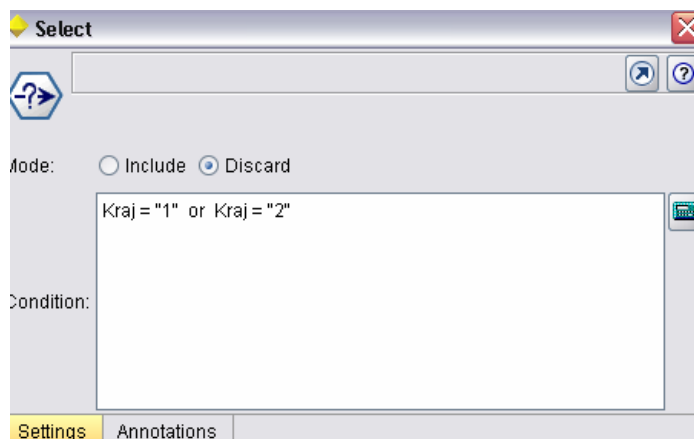
V této části bude provedena korekce vstupních dat pomocí uzlu „Reclassify” a „Select”. Jelikož původní data u atributu Kraj představovala označení kraje příslušným číslem, byla provedena korekce v podobě přiřazení číselné hodnotě textový název. Jedná se o úpravu, která nám nijak neovlivní výsledek analýzy, slouží nám pouze k lepší orientaci v datovém souboru.

Obrázek 13 - Reklasifikace atributu Kraj [Zdroj: vlastní]



V dalším kroku provedeme pomocí uzlu „Select” výběr z atributu Kraj. Bude se jednat o data týkající se České republiky a hlavního města Prahy. Tyto údaje budou odstraněny z důvodu jejich extrémních hodnot, které by zkreslovaly výslednou analýzu.

Obrázek 14 - Vyřazení atributů [Zdroj: vlastní]

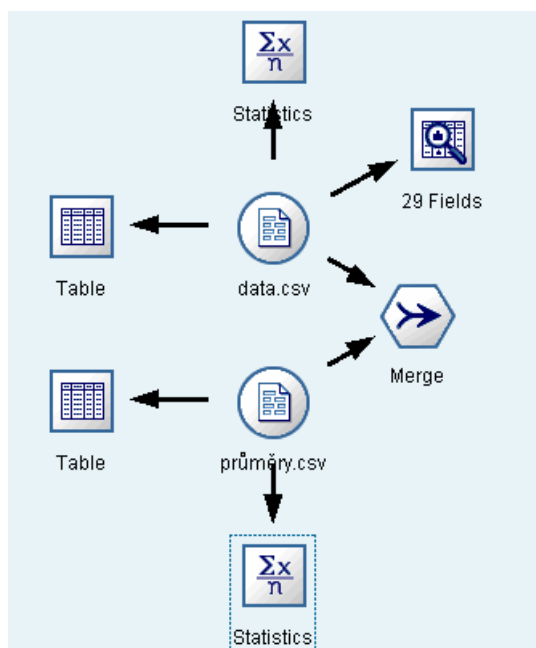


3.3 Přepočítání hodnot na 10 000 obyvatel

Z důvodu různého počtu obyvatel v jednotlivých krajích bylo nutné tyto hodnoty přepočítat. Tento krok byl proveden v programu Clementine. Nejdříve byl však v programu MS Excel proveden přepočítání na průměrnou hodnotu za roky 1998 – 2007. Tento krok byl proveden z důvodu toho, že by se pokaždé kraj přiřadil do jiné zdravotní úrovně v různém roce. Mohlo by se tedy stát, že všechny kraje by např. v roce 1998 spadaly do nejnižší úrovně a naopak v roce 2007 do té nejvyšší. V programu Clementine by tento přepočítání musel probíhat u každého kraje zvlášť a obsahoval by mnoho uzlů, které by nám zhoršovaly orientaci v modelu. Dále, abychom tyto kraje mezi sebou mohli porovnávat, byly hodnoty přepočítány na 10 000 obyvatel. Přepočítání na obyvatele bylo provedeno v programu Clementine za pomoci uzlu „Derive“. V obou případech se jednalo pouze o atributy z oblasti zdravotnictví.

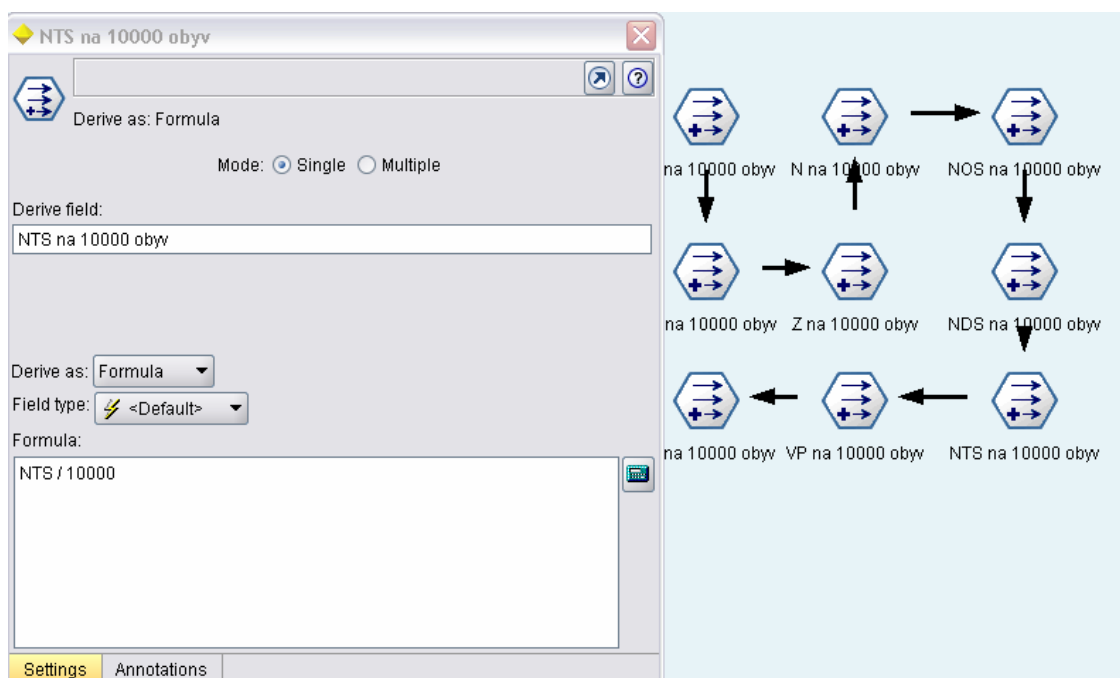
V prvním kroku byl tedy vytvořen další datový soubor s jednotlivými průměrnými hodnotami za jednotlivé roky. Použit byl pro nahrání dat uzel „Var. File“, jeho vstupní hodnoty a datový slovník jsou obsahem Přílohy č. 5 a 6. Propojení s původním datovým souborem bylo provedeno pomocí uzlu „Merge“ a jako klíč byl zvolen atribut Kraj. Ukázka je zobrazena na Obrázku č. 15.

Obrázek 15 - Propojení dvou datových souborů [Zdroj: vlastní]



V druhém kroku byly přepočítány pomocí uzlu „Derive“ hodnoty týkající se zdravotnictví na 10 000 obyvatel.

Obrázek 16 - Přepočet na 10 000 obyvatel [Zdroj: vlastní]



3.4 Rozdělení vstupních dat na trénovací a testovací množinu

V závěrečné části korekce je dobré data rozdělit na trénovací a testovací množinu. Tento krok bude proveden pomocí uzlu „Partition“. Trénovací data by měla tvořit zhruba dvě třetiny původního datového souboru a testovací data jednu třetinu. Je k dispozici také možnost, kdy jsou data rozdělena na trénovací, testovací a validační. V tomto případě postačí první varianta, jelikož máme relativně malý objem dat.

Datový soubor byl rozdělen v poměru 70% ku 30%, což je zřetelné z následujícího obrázku. V případě použití různých metod modelování musí být zajištěno, že rozdělení datového souboru bude vždy stejné. To opatří zaškrtnuté políčko „Set random seed“.

Obrázek 17 - Rozdělení dat na trénovací a testovací množinu - MATRIX [Zdroj: vlastní]

Zdrav.úroveň						
Partition		nadprůměr...	nízká	podprůměr...	průměrná	vyšoká
1_Training	Count	13	7	9	14	47
	Total %	10.000	5.385	6.923	10.769	36.154
2_Testing	Count	7	3	1	6	23
	Total %	5.385	2.308	0.769	4.615	17.692

4 NÁVRH MODELU

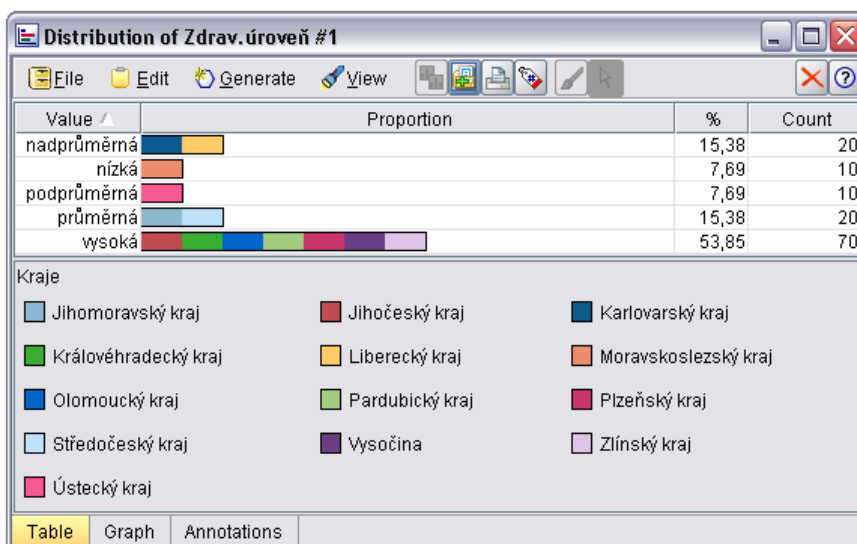
Tato kapitola se věnuje modelování dat, které bude provedeno za pomoci programu Clementine. Pro modelování budou použity rozhodovací stromy a shluková analýza. V první části kapitoly bude provedena analýza, zda životní prostředí v kraji ovlivňuje zdravotní stav obyvatel. V druhé části se zaměříme na to, zda zdravotní stav obyvatel ovlivňuje jejich pracovní výkonnost a tím i finanční situaci kraje. V závěru kapitoly bude kvalita života odhadována pomocí vybraných ukazatelů. Na základě těchto výsledků bude zachycen návrh modelu.

4.1 Vliv životního prostředí na zdravotní stav obyvatel

V úvodní části práce byly kraje rozděleny podle jejich zdravotní úrovně do shluků pomocí shlukové analýzy. K tomu byl využit uzel K-Means, kde byl přednastaven maximální počet shluků 5. Tento počet shluků byl volen z důvodu snížení množství dat v důsledku zprůměrnování za jednotlivé roky. Tyto shluky byly dále reklasifikovány pomocí uzlu „Derive“ na shluky krajů s vysokou, nadprůměrnou, průměrnou, podprůměrnou a nízkou životní úrovní. Podrobnější informace obsahuje Příloha č. 7.

Jak je vidět z následujícího grafu „Distribution“, do shluku 2 a 3 byl zařazen pouze jeden kraj a to kraj Ústecký (podprůměrná) a Moravskoslezský (nízká). Po dvou krajích obsahují shluky 4 a 5, což jsou kraje s nadprůměrnou a průměrnou životní úrovní. Do nejvyšší úrovně spadá zbývajících 7 krajů.

Obrázek 18 - Rozdělení do shluků - zdravotní úroveň [Zdroj: vlastní]



4.1.1 Rozhodovací stromy C&RT a CHAID

U rozhodovacích stromů bylo nutné nadefinovat, co bude vstupem a výstupem, jelikož se jedná o analytické nástroje učení s učitelem.

Obrázek 19 - Vstupy a výstupy rozhodovacího stromu - zdravotní úroveň [Zdroj: vlastní]

Field	Type	Values	Missing	Check	Direction
Rok	Ordered Set	1998,1999...		None	None
Naděje na d...	Range	[68.8,75.6]		None	None
Naděje na d...	Range	[76.1,81.1]		None	None
Počet nemoc...	Range	[5,225]		None	None
Počet lékařů	Range	[1021,436...		None	None
PIO ŽP	Range	[227951,4...		None	In
Tuhé látky	Range	[535.7,756...		None	In
SO2	Range	[1422.3,43...		None	In
NOx	Range	[1751.6,16...		None	In
CO	Range	[2117.8,39...		None	In
ŽN	Range	[2991.4,96...		None	None
MN	Range	[11.4,281.8]		None	None
Z	Range	[2975.6,10...		None	None
N	Range	[859.4,285...		None	None
NOS	Range	[1387.3,56...		None	None
NDS	Range	[166.6,505...		None	None
NTS	Range	[136.0,449...		None	None
VP	Range	[201.1,673...		None	None
PL	Range	[1083.7,39...		None	None
Kraje	Set	"Hl. město...		None	None
ŽN na 10000...	Range	[0.29914,1...		None	None
MN na 10000...	Range	[0.00114,0...		None	None
Z na 10000 o...	Range	[0.29756,1...		None	None
N na 10000 ...	Range	[0.08594,0...		None	None
NOS na 100...	Range	[0.13873,0...		None	None
NDS na 100...	Range	[0.016659...		None	None
NTS na 1000...	Range	[0.0136,0...		None	None
VP na 10000 ...	Range	[0.02011,0...		None	None
PL na 10000 ...	Range	[0.108370...		None	None
\$KM-K-Means	Set	cluster-1,c...		None	None
Zdrav.úroveň	Ordered Set	nadprůmě...		None	Out
Partition	Set	"1_Trainin...		None	Partition

Rozhodovací strom C&RT se využívá v problémových případech, kdy máme vysoká čísla u vstupních dat nebo nám některá data chybí. Pracuje s kategorizovanými i spojitými proměnnými, kdy vstupní proměnná může být jedna a více proměnných, avšak na výstupu vyžaduje pouze jednu proměnnou, která může být jak kategorizovaná, tak spojitá. Jedná se o binární stromový algoritmus. [25]

Výstupem metody byl „Generate model“, neboli obecný model a maximální počet větvení pod kořenem stromu bylo nastaveno na 5. Dále bylo zaškrtnuto políčko „Use partitioned data“, aby bylo zachováno rozdělení dat na trénovací a testovací množinu.

Rozhodovací strom CHAID je rychlý statistický algoritmus, který se využívá jako kritérium pro větvení χ^2 . Štěpí skupiny vždy na vhodný počet statisticky homogenních podskupin. Při větvení nevytváří tolik větví, kolik má atribut hodnot a hodnoty atributu postupně seskupuje z původního počtu až do dvou skupin. Výsledkem je obecný strom, kdy z každého uzlu vychází maximálně tolik větví, kolik je počet kategorií proměnné, která slouží jako prediktor. [25]

Na základě výsledků analýzy, byly jako výstup metody zvoleny výše uvedené rozhodovací stromy, způsob tvorby postačí jednoduchý, je preferována přesnost před všeobecností a v poslední řadě je důležité zaškrtnutí pole „Use partitioned data“, tedy rozdělení dat na trénování a testovací množinu. Výstup obou stromů je v Příloze 8 a 9.

4.1.2 Porovnání výsledků vybraných stromů

Na následujícím obrázku je zachycen výsledek predikce vybraných stromů:

Obrázek 20- Analýza výsledků C&RT a CHAID [Zdroj: vlastní]

Results for output field Zdrav.úroveň

Individual Models

Comparing \$R-Zdrav.úroveň with Zdrav.úroveň

'Partition'	1_Training		2_Testing	
Correct	90	100%	37	92,5%
Wrong	0	0%	3	7,5%
Total	90		40	

Comparing \$R1-Zdrav.úroveň with Zdrav.úroveň

'Partition'	1_Training		2_Testing	
Correct	87	96,67%	38	95%
Wrong	3	3,33%	2	5%
Total	90		40	

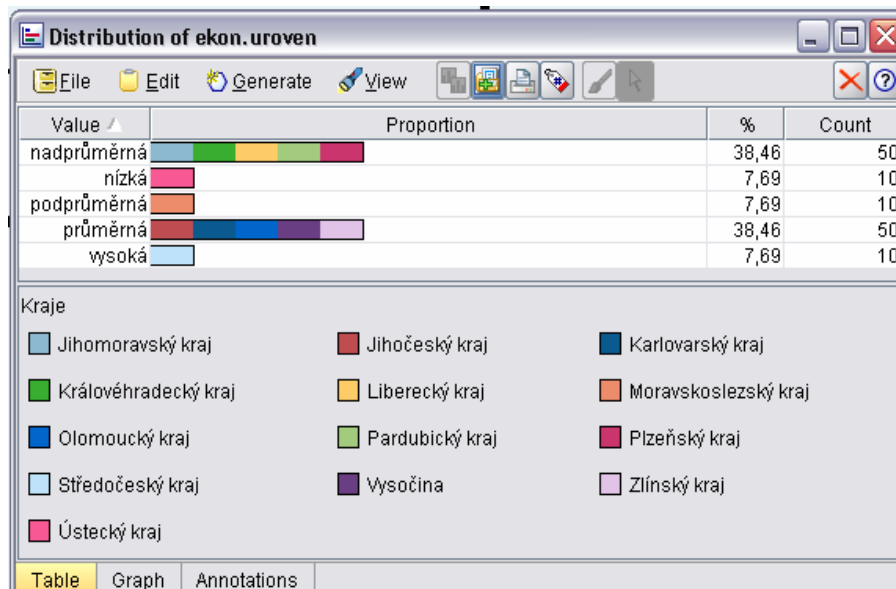
Výsledek zjištěný pomocí uzlu „Analysis“ zachycuje, že rozhodovací stromy C&RT a CHAID predikovaly s nejvyšší přesností. Na trénovacích datech dokonce rozhodovací strom C&RT nechyboval vůbec a na testovacích pouze v 7,5 %. U stromu CHAID proběhl odhad ze 3,33 % špatně na trénovacích datech a v 5 % na testovacích datech. U ostatních stromů byly větší odchylky v predikci a proto nebyly dále do modelu zahrnuty. Výsledný návrh modelu je zachycen v závěru této kapitoly.

4.2 Vliv zdravotního stavu obyvatel na ekonomiku kraje

Stejně jako v předcházející části, byly kraje rozděleny do shluků pomocí uzlu „K-Means“ na kraje s vysokou, nadprůměrnou, průměrnou, podprůměrnou a nízkou ekonomickou úrovní. Jak lze odvodit, jednalo se o rozdělení do pěti shluků, které byly posléze přejmenovány pomocí uzlu „Derive“ na výše uvedené úrovně. Toto rozřídění probíhalo na základě ukazatelů hrubého domácího produktu, čistého disponibilního důchodu, míry registrované nezaměstnanosti a hrubé mzdy v jednotlivých krajích. Podrobnější informace jsou zachyceny v Příloze č. 10.

Na distribučním grafu je patrné rozdělení krajů do jednotlivých úrovní ekonomického rozvoje. Tři kraje nejsou podobné ani jednomu ze souboru a tak byly pomocí shlukové analýzy rozděleny každý do jednoho shluku. Ústecký kraj spadl do nejnižší ekonomické úrovně, Moravskoslezský do podprůměrné a Středočeský kraj spadá naopak do nejvyšší. Další kraje se dělí o úroveň nadprůměrnou a průměrnou. Do nich patří většina krajů z datového souboru. Z grafu je možné vidět, že obě úrovně obsahují 38,46% dat ze souboru. Na základě toho lze také vyvodit, že většina krajů České republiky má dobrou ekonomickou výkonnost.

Obrázek 21 - Rozdělení do shluku - ekonomická úroveň [Zdroj: vlastní]



Distribuční graf byl získán za použití uzlu „Distribution“.

4.2.1 Rozhodovací stromy CHAID, QUEST a C&RT

Jelikož bylo zjišťováno, zda ekonomickou úroveň kraje ovlivňují zdravotní indikátory v daném kraji, byly jako vstupy zvoleny následující ukazatele. Jako vstupní hodnoty byla použita data, která jsme si přepočítali na 10 000 obyvatel a za výstupní hodnotu je zvolena množina krajů rozdělená podle jejich ekonomické úrovně. Tento výstup byl získán na základě shlukové analýzy.

Obrázek 22 - Vstupy a výstupy rozhodovacího stromu [Zdroj: vlastní]

Field	Type	Values	Missing	Check	Direction
Rok	Ordered Set	1998,1999...		None	None
Kraje	Set	"Hl. město...		None	None
ŽN na 10000...	Range	[0.29914,1...		None	In
MN na 10000...	Range	[0.00114,0...		None	In
Z na 10000 o...	Range	[0.29756,1...		None	In
N na 10000 ...	Range	[0.08594,0...		None	In
NOS na 100...	Range	[0.13873,0...		None	In
NDS na 100...	Range	[0.016659...		None	In
NTS na 1000...	Range	[0.0136,0...		None	In
VP na 10000 ...	Range	[0.02011,0...		None	In
PL na 10000 ...	Range	[0.108370...		None	In
Zdrav.úroveň	Ordered Set	nadprůmě...		None	None
\$KM1-K-Mea...	Set	cluster-1,c...		None	None
ekon.uroveň	Ordered Set	nadprůmě...		None	Out
Partition	Set	"1_Trainin...		None	Partition

Nejlépe výsledků dosáhly rozhodovací stromy CHAID, QUEST a C&RT, které předikovaly se 100% přesností. Těchto výsledků bylo dosaženo zejména na základě výběru dat, které máme k dispozici v datovém souboru.

Rozhodovací strom QUEST je statistický algoritmus pro selekci proměnných bez vychýlení, který rychle a účinně sestavuje binární stromy. Vybírá proměnné a štěpí jejich hodnoty do skupin v oddělených krocích, aby zvýšil přesnost a rychlost provedení. [25]

Stejně jako v předchozím případě byl zvolen způsob tvorby jednoduchý, byla preferována přesnost před všeobecností a bylo zaškrtnuto pole „Use partitioned data“, tedy zachování rozdělení dat na trénovací a testovací množinu. Výstupy nejlépe predikovaných stromů jsou v Přílohách 11, 12 a 13, kde je zobrazeno grafické vyjádření buď celé nebo určité části rozhodovacích stromů.

4.2.2 Porovnání výsledků vybraných stromů

Na následujícím obrázku je zachycen výsledek predikce vybraných stromů:

Obrázek 23 - Analýza výsledků CHAID, QUEST a C&RT [Zdroj: vlastní]

Results for output field ekon.uroveň

Individual Models

Comparing \$R-ekon.uroveň with ekon.uroveň

'Partition'	1_Training		2_Testing	
Correct	90	100%	40	100%
Wrong	0	0%	0	0%
Total	90		40	

Comparing \$R1-ekon.uroveň with ekon.uroveň

'Partition'	1_Training		2_Testing	
Correct	90	100%	40	100%
Wrong	0	0%	0	0%
Total	90		40	

Comparing \$R2-ekon.uroveň with ekon.uroveň

'Partition'	1_Training		2_Testing	
Correct	90	100%	40	100%
Wrong	0	0%	0	0%
Total	90		40	

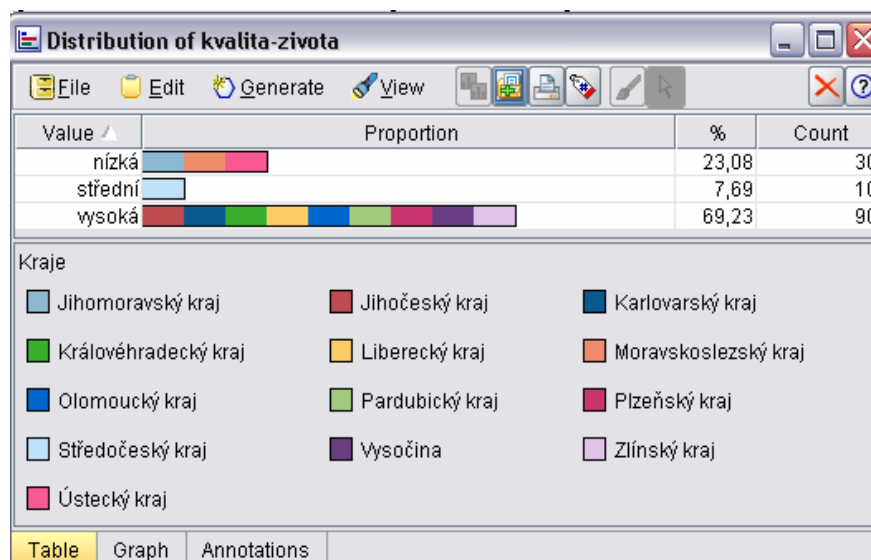
Tento výsledek byl zachycen na základě uzlu „Analysis“ a z výše uvedené analýzy lze zaznamenat nulovou chybovost většiny analytických nástrojů, které byly pro predikci použity.

4.3 Kvalita života zjištěná na základě vybraných ukazatelů

Datový soubor obsahoval mimo ekonomických a zdravotních ukazatelů také doplňující ukazatele, které byly v této části využity na posouzení kvality života jedinců v daném kraji. Jednalo se o ukazatele celkového přírůstku obyvatel v daném kraji, obecnou kriminalitu, rozvodovost a sňatečnost v kraji. Ukazatel celkového přírůstku obyvatel v tomto případě znamenal spokojenost lidí s danou lokalitou. Jinými slovy, jestliže jsou lidé v daném kraji spokojení, zvyšují počet obyvatel například v důsledku jejich přistěhování nebo statistiku porodnosti, apod.. V opačném případě, pokud by lidé nechtěli v daném kraji do budoucna setrávat, nebudou plánovat rodinu či jiné záležitosti. Za prvé z důvodu například nedostatku financí, které další člen rodiny, či změna bydliště vyžaduje a na straně druhé v důsledku dalších změn, které stěhování přináší jako např. změna základní školy či mateřské školky, atd.. Se spokojeností v kraji souvisí také další ukazatel, kterým je obecná kriminalita v kraji. V případě, že se v kraji vyskytuje vysoká míra zjištěných trestných činů, je možné, že se lidé budou soustředit z důvodu bezpečnosti jinam. Rozvodovost a sňatečnost jsou poslední indikátory přispívající k obecné spokojenosti obyvatel.

Na základě těchto ukazatelů byly kraje rozděleny do tří shluků za pomoci uzlu „K-Means“, které zahrnovaly kraje s vysokou, střední a nízkou kvalitou života. Toto členění bylo provedeno uzlem „Derive“, který každému shluku přiřadil textové vyjádření. Rozdělení do pěti shluků nebylo realizováno z důvodu vyšší podobnosti krajů vzhledem k hodnotám těchto ukazatelů. Rozdělení dle kvality života je zobrazeno na následujícím obrázku.

Obrázek 24 - Rozdělení do shluků - obecná kvalita života [Zdroj: vlastní]



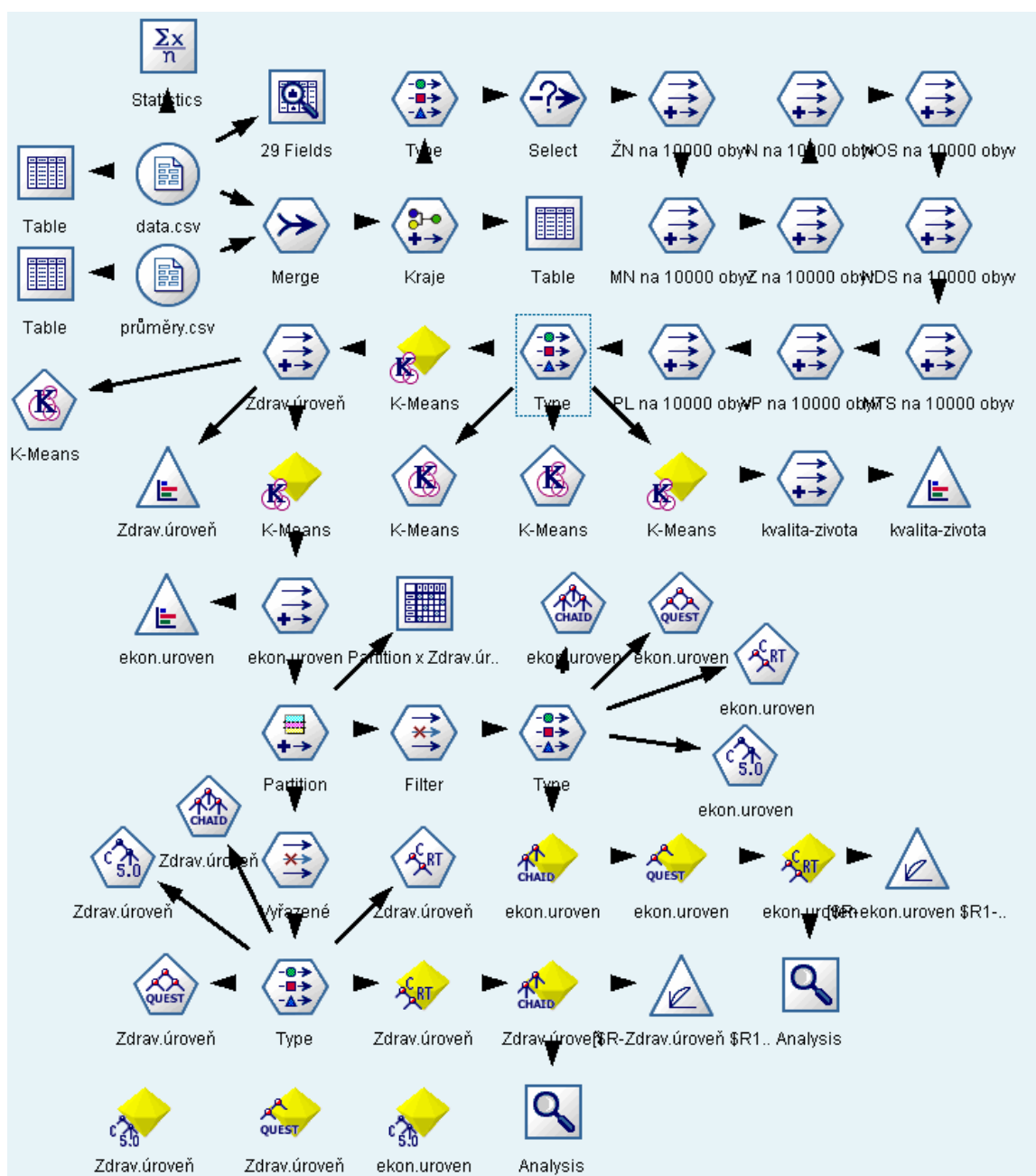
Z distribučního grafu lze vysledovat, že do úrovně vysoké kvality života, definované vybranými ukazateli, spadá nejvíce krajů České republiky, do střední úrovně patří pouze kraj Středočeský a nejhorší kvalitu života, na základě výše uvedených ukazatelů, mají kraje Jihomoravský, Ústecký a Moravskoslezský.

Celkové rozdělení krajů do shluků je zachyceno v Příloze č. 14.

4.4 Výsledný STREAM

V této části je zobrazen návrh samotného modelu v programovém prostředí Clementine. Na obrázku v dolní části jsou vyřazeny 3 rozhodovací stromy, které nejsou součástí modelu, jelikož jejich výsledky byly nejhorší. V rámci upravování nebyly však úplně smazány, aby se mohly kdykoliv připojit a bylo možné se na tyto závěry opětovně podívat.

Obrázek 25 - Výsledný STREAM [Zdroj: vlastní]



5 ANALÝZA VÝSLEDKŮ

Pro predikci kvality života bylo využito analytických nástrojů učení s učitelem, jehož hodnocení modelu je založeno na testování shody nalezených hodnot s informací učitele. Shoda nalezených hodnot s informací učitele se v tomto případě zjišťovala za pomoci uzlu „Analysis“, který zobrazuje počet správně a špatně určených hodnot. Tyto informace lze zachytit i na základě grafického vyjádření a to při použití uzlu „Evaluation“

Na následujícím obrázku jsou vidět analýzy všech použitých rozhodovacích stromů. V tomto případě se jedná o analýzu kvality života zdraví ovlivněnou životním prostředím.

Obrázek 26 - Analýza odhadu rozhodovacích stromů [Zdroj: vlastní]

Results for output field Zdrav.úroveň

Individual Models

- Comparing \$R-Zdrav.úroveň with Zdrav.úroveň

'Partition'	1_Training	2_Testing
Correct	90 100%	37 92,5%
Wrong	0 0%	3 7,5%
Total	90	40
- Comparing \$C-Zdrav.úroveň with Zdrav.úroveň

'Partition'	1_Training	2_Testing
Correct	89 98,89%	35 87,5%
Wrong	1 1,11%	5 12,5%
Total	90	40
- Comparing \$R1-Zdrav.úroveň with Zdrav.úroveň

'Partition'	1_Training	2_Testing
Correct	72 80%	33 82,5%
Wrong	18 20%	7 17,5%
Total	90	40
- Comparing \$R2-Zdrav.úroveň with Zdrav.úroveň

'Partition'	1_Training	2_Testing
Correct	87 96,67%	38 95%
Wrong	3 3,33%	2 5%
Total	90	40
- Agreement between \$R-Zdrav.úroveň \$C-Zdrav.úroveň \$R1-Zdrav.úroveň \$R2-Zdrav.úroveň

'Partition'	1_Training	2_Testing
Agree	68 75,56%	31 77,5%
Disagree	22 24,44%	9 22,5%
Total	90	40
- Comparing Agreement with Zdrav.úroveň

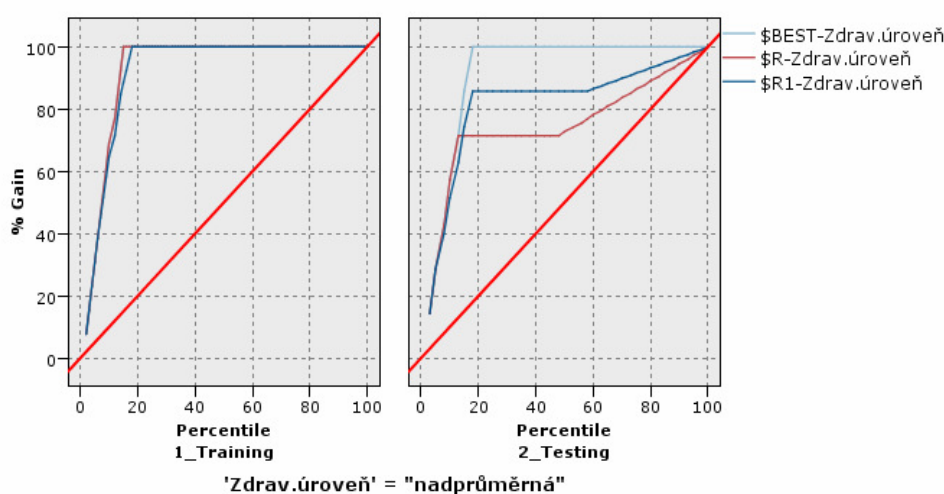
'Partition'	1_Training	2_Testing
Correct	68 100%	31 100%
Wrong	0 0%	0 0%
Total	68	31

Z Obrázku č. 26 je patrné, že nejlepších výsledků dosáhl na trénovacích datech algoritmus C&RT, který predikoval s 100% úspěšností. Nejvyšší odhad na testovacích datech měl algoritmus CHAID, který predikoval s 95% úspěšností.

Poslední dvě tabulky ukazují, v kolika případech se metody shodly. U testovacích dat je tato shoda pouze na 77.5%, avšak z toho je správných 100%.

Průběh učení je zachycen na následujícím obrázku, kdy obě metody predikovaly s velmi vysokou přesností. Nejlepších výsledků dosáhl analytický nástroj C&RT, který je znázorněn červenou barvou a algoritmus CHAID, který měl větší odchylky v celkové predikci, je zobrazen modře. Světle modrá křivka představuje nejlepší možnou variantu odhadu. Rozhodovací stromy z obou metod jsou v Příloze č. 8 a 9.

Obrázek 27 - Evaluační graf - srovnání odhadu C&RT a CHAID [Zdroj: vlastní]



V další části práce byla provedena analýza odhadu kvality života z hlediska ekonomického. Kvalita života byla v tomto případě ovlivňována zdravotním stavem obyvatel. Jinými slovy bylo zjišťováno, zda kraje s vyšší zdravotní úrovní a tím i větší možností pracovních sil, pokud není brána v úvahu dobrovolná nezaměstnanost, mají vliv na jeho ekonomiku.

Výsledná analýza je zachycena na Obrázku č. 28. Predikce zde dopadla mnohem lépe. Jak je vidět z posledních dvou tabulek, shody dosáhly jednotlivé metody v 92.5% na testovacích datech a v 92.22% na trénovacích datech, z toho správně odhadnuté hodnoty byly všechny.

Obrázek 28 - Analýza odhadu rozhodovacích stromů [Zdroj: vlastní]

Results for output field ekon.uroven

Individual Models

- Comparing \$R-ekon.uroven with ekon.uroven

'Partition'	1_Training		2_Testing	
Correct	90	100%	40	100%
Wrong	0	0%	0	0%
Total	90		40	
- Comparing \$R1-ekon.uroven with ekon.uroven

'Partition'	1_Training		2_Testing	
Correct	90	100%	40	100%
Wrong	0	0%	0	0%
Total	90		40	
- Comparing \$R2-ekon.uroven with ekon.uroven

'Partition'	1_Training		2_Testing	
Correct	90	100%	40	100%
Wrong	0	0%	0	0%
Total	90		40	
- Comparing \$C-ekon.uroven with ekon.uroven

'Partition'	1_Training		2_Testing	
Correct	83	92,22%	37	92,5%
Wrong	7	7,78%	3	7,5%
Total	90		40	

Agreement between \$R-ekon.uroven \$R1-ekon.uroven \$R2-ekon.uroven \$C-ekon.uroven

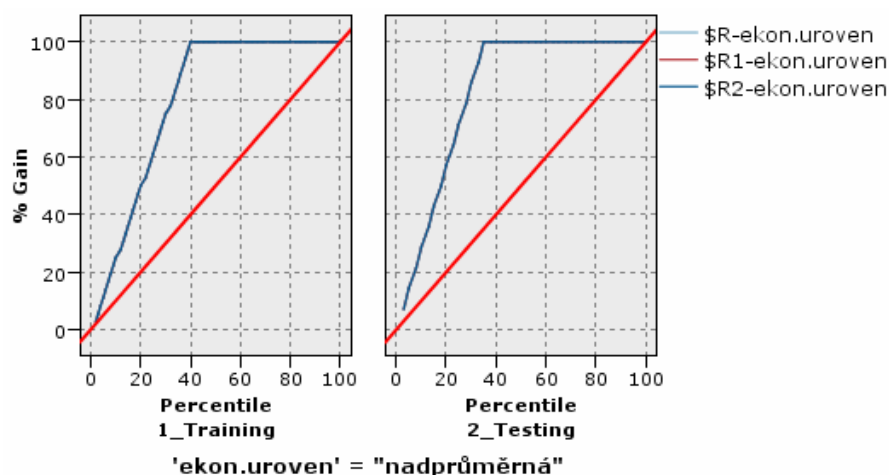
'Partition'	1_Training		2_Testing	
Agree	83	92,22%	37	92,5%
Disagree	7	7,78%	3	7,5%
Total	90		40	

Comparing Agreement with ekon.uroven

'Partition'	1_Training		2_Testing	
Correct	83	100%	37	100%
Wrong	0	0%	0	0%
Total	83		37	

Z evaluačního grafu je vidět grafické znázornění předchozí analýzy. Na obrázku jsou zachyceny pouze stromy předikující se 100% úspěšností. Světle modrá barva představuje stromový diagram CHAID, červená statistický algoritmus QUEST a modrá stromový algoritmus C&RT. Na grafu je vidět pouze poslední z algoritmů, jelikož jejich odhady byly totožné a tudíž dochází k překrývání jejich výsledků.

Obrázek 29 - Evaluační graf - srovnání odhadu CHAID, QUEST a C&RT [Zdroj: vlastní]



6 ZÁVĚR

V práci jsem se zabývala analýzou přístupů k modelování kvality života. Tento problém byl řešen za pomoci shlukové analýzy a rozhodovacích stromů. K řešení příkladů a tvorbě ukázek byl volen programový prostředek vytvořený firmou SPSS Clementine 11.1.

Na základě ukazatelů, týkajících se různých oblastí našeho života, byla zjištěna kvalita života v jednotlivých krajích. Nejprve byl zkoumán vliv životního prostředí na zdraví jedince, za pomoci kterého bylo zjištěno, že do nejvyšší kvality života spadá většina krajů České republiky. Nadprůměrnou kvalitu života mají poté v kraji Karlovarském a Libereckém, průměrnou v kraji Středočeském a Jihomoravském. Na předposlední úrovni se nachází kraj Ústecký a nejnižší úroveň reprezentuje kraj Moravskoslezský.

V další části byl zkoumán vliv zdravotního stavu obyvatel na ekonomickou výkonnost kraje. Z analýz vyšel nejlépe kraj Středočeský, o stupeň níže se umístil kraj Plzeňský, Pardubický, Liberecký, Královéhradecký, Jihomoravský. Jako průměrný se umístil kraj Zlínský, Vysočina, Olomoucký, Karlovarský a Jihočeský. Na posledních místech se jako v předchozí analýze nachází kraj Ústecký a Moravskoslezský, zde ovšem v opačném pořadí. Nejhorší ekonomickou úroveň má tedy Ústecký kraj.

V poslední analýze dopadly výsledky téměř stejně jako v předchozích dvou analýzách. Výjimku tvořil snad jen Jihomoravský kraj, který se na základě zvolených ukazatelů zařadil na nejnižší úroveň. Tato analýza ale probíhala na základě 4 atributů, tudíž je více než možné, že by tento kraj, v případě více ukazatelů, spadl do úrovně vyšší.

Ze všech tří analýz lze do krajů s vysokou kvalitou života zařadit kraj Plzeňský, Královéhradecký, Pardubický a Zlínský. Do krajů s nadprůměrnou kvalitou života lze zahrnout kraj Jihočeský, Karlovarský, Liberecký, Vysočinu a kraj Olomoucký. Průměrnou kvalitu života mají v kraji Středočeském a Jihomoravském, který stojí na pomezí mezi průměrem a podprůměrem. Nejhorší kvalitu života mají v kraji Ústeckém a Moravskoslezském. Tento fakt může být způsobený např. útlumem v těžkém průmyslu, který výrazně zvyšuje nezaměstnanost v krajích a nechutí obyvatel orientovat se na tuto oblast. To je také zřejmé z výsledků analýzy celkového přírůstku obyvatel, kdy Moravskoslezský kraj disponuje nejvyšším počtem obyvatel ze všech českých krajů a celkový

přírůstek obyvatel měl ze všech krajů nejhorší. Je zde také velký problém se špatnou kvalitou životního prostředí, který může být také připisován těžkému průmyslu.

Relativně špatný výsledek Středočeského kraje lze přisuzovat jeho těsné blízkosti k hlavnímu městu. Stejně tak mu lze na druhou stranu přičíst jeho vysokou ekonomickou výkonnost. Na tomto faktu má ale také zejména zásluhu největší český vývozce Škoda Mladá Boleslav, která zaměstnává okolo 25 000 lidí. U jihomoravského kraje může být tento nevalný výsledek spojován se zhoršenou situací zemědělství v České republice, na kterou se tento kraj převážně orientuje.

Diplomová práce potvrdila skutečnost, že kvalita života jedince, může být predikována na základě výše vybraných ukazatelů. Cíl práce byl splněn.

SYMBOLY A ZKRATKY

ČSÚ – Český statistický úřad

EQLS - European quality of life survey (průzkum kvality života v Evropě)

GDI – Gender-related Development Index (index lidského rozvoje zohledňující nerovnost mezi pohlavími)

GEM - Gender Empowerment Measure (index měřící aktivní účast žen na ekonomickém a politickém životě)

HDI – Human Development Index (Index lidského rozvoje)

HPI – Human Povarty Index (index lidské chudoby)

HDP - Hrubý domácí produkt

REZZO – Registr emisí a zdrojů znečišťování ovzduší

RISY – Portál regionálních informačních servisů

ŠVOK – Študentská vedecká a odborná konferencia

TDIDT – Top down induction of decision trees

UNDP - United Nations Development Programme

USD - United States dollar - americký dolar

ÚZIS – Ústav zdravotnických informací a statistiky ČR

POUŽITÁ LITERATURA

- [1] Berka, P. *Dobývání znalostí z databází*. Praha: Academia, 2003. ISBN 80-200-1062-9.
- [2] *Český statistický úřad* [online]. 2008 [cit. 2010-03-13]. Naděje dožití a průměrný věk. Dostupné z WWW: <http://www.czso.cz/csu/redakce.nsf/i/nadeje_dozeni_a_prumerny vek>.
- [3] *Český statistický úřad* [online]. 2009 [cit. 2010-02-09]. Dostupný z WWW: <<http://www.czso.cz/csu/2009edicniplan.nsf/kapitola/0001-09-2009-0100>>.
- [4] *Český statistický úřad* [online]. 2010 [cit. 2010-02-09]. Dostupný z WWW: <http://www.czso.cz/csu/redakce.nsf/i/rocenky_souhrn>.
- [5] Džuka, J.: *Psychologické dimenzie kvality života*. Prešov: Prešovská univerzita, 2005. ISBN 80-8068-282-8.
- [6] *Grafologie a psychologie* [online]. 2008 [cit. 2009-11-11]. Dostupný z WWW: <<http://ografologii.blogspot.com/2008/04/abraham-maslow-pyramida-potreb.html>>.
- [7] Han, J., Kamber, M.: *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufman Publishers, 2001. ISBN 1-55860-901-6.
- [8] *Human Development Reports* [online]. 2007 [cit. 2009-11-24]. Dostupný z WWW: <<http://hdr.undp.org/en/statistics>>.
- [9] Chapman, A. Businessballs.com [online]. 1995-2009 [cit. 2009-12-01]. Dostupný z WWW: <<http://www.businessballs.com/maslow.htm>>.
- [10] Kašparová, M., Křupka, J., Pírko, J.: *Modelování spokojenosti občanů ve vztahu k regionálnímu rozvoji a kvalitě života*. Scientific Papers ? Series D, Univerzita Pardubice, roč. 13, 2008, s. 109-120. ISBN 978-80-7395-040-8, ISSN 1211-555X.
- [11] *Kvalita života – různé koncepty* [online]. 2009 [cit. 2009-11-20]. Dostupný z WWW: <holas.xf.cz/vse/5HP400.doc>.
- [12] Křivohlavý, J.: *Psychologická pojetí a způsoby zjišťování kvality života*. In: *Psychologické aspekty kvality života*. Sborník příspěvků z mezinárodního a celostátního kola ŠVOK. Bratislava, 2001, s. 20-41.
- [13] Křivohlavý, J.: *Psychologie nemoci*. Praha: Grada Publishing, 2002. ISBN 80-247-0179-0.
- [14] Křivohlavý, J.: *Kvalita života*. In: *Kvalita života*. Sborník příspěvků z konference. Třeboň, 2004, s. 9-20.
- [15] Machovec, M. 2002: *Smysl lidské existence*. Praha: Akropolis, ISBN 80-7304-016-6.
- [16] *Metody shlukové analýzy* [online]. 2005-2006 [cit. 2010-03-06]. Dostupné z WWW: <http://www2.zf.jcu.cz/public/departments/kmi/MSMT_05/metody%20shlukove%20analyzy.pdf>.

- [17] Payne, J. a kol. 2005: *Kvalita života a zdraví*. Praha: Triton, ISBN 80-7254-657-0.
- [18] Petr, P.: *Data Mining Díl I*. Pardubice: Univerzita Pardubice, 2008. ISBN 80-7395-098-9.
- [19] Phillips, D.: *Quality of Life : Concept, Policy and Practice*, 2006. ISBN 978-0-415-32355-0.
- [20] Potůček, J.: *Metodologie modelování biologických systémů*. Praha: Česká technika – nakladatelství ČVUT, 2009. ISBN 80-01-04412-4.
- [21] *Portál regionálních informačních servisů* [online]. 2010 Dostupné z WWW: <<http://www.risy.cz>>.
- [22] Rapley, M.: *Quality of Life Research*, 2003. ISBN 978-0-7619-5457-6.
- [23] *Rozhodovací stromy* [online]. 2002 [cit. 2009-03-20]. Dostupný z WWW: <<http://datamining.xf.cz/view.php?cislocianku=2002102802>>.
- [24] SPSS. *SPSS Classification Trees* [online]. 2008 [cit. 2009-03-13]. Dostupné z: <http://www.spss.cz/sw_mcla.htm>.
- [25] *Segmenty, asociační interakce, profily a struktura dat* [online]. 2008 [cit. 2009-04-15]. Dostupné z: <http://www.spss.cz/files/programy/sw_answertree.pdf>.
- [26] *Soudobá sociologie II - Kvalita života* [online]. 2009 [cit. 2009-11-20]. Dostupné z WWW: <kvalitazivota.vubp.cz/.../soudoba_sociologie_II_kvalita_zivota-svobodova.doc>.
- [27] *Systém monitorování zdravotního stavu obyvatelstva České republiky ve vztahu k životnímu prostředí* [online]. 2009 [cit. 2010-03-06]. Dostupné z WWW: <http://www.szu.cz/uploads/documents/chzp/souhrnna_zprava/Szu_09_1_1.pdf>.
- [28] Šarmanová, J. *Metody dolování znalostí z dat* [online]. 2002 [cit. 2010-03-06]. Dostupné z WWW: <http://www.datakon.cz/datakon08/d02_sarmanova.pdf>.
- [29] Ústav ochrany čistoty ovzduší [online]. 2000 – 2007. Dostupný z WWW. <http://www.chmi.cz/uoco/oco_main.html>.
- [30] Ústav zdravotnických informací a statistiky ČR [online]. 2007. Dostupný z WWW: <http://www.uzis.cz/news.php?mnu_id=1100>
- [31] Vavroušek J.: *Modelování biologických a sociálních objektů*. Praha: Academia, 1989. ISBN: 80-200-009-2.
- [32] *Veřejná databáze ČSÚ* [online]. 2010 [cit.2009-03-13]. Dostupný z WWW: <http://vdb.czso.cz/vdbvo/maklist.jsp?kapitola_id=2&expand=1>.
- [33] *Vliv změn světa práce na kvalitu života* [online]. 2007 [cit. 2010-02-09]. Dostupný z WWW: <http://kvalitazivota.vubp.cz/prispevky/svet_prace.pdf>.
- [34] *Wikipedia* [online]. 2009 [cit. 2009-11-11]. Dostupný z WWW: <<http://cs.wikipedia.org>>.

PŘÍLOHY

PŘÍLOHA č. 1: UKÁZKA VSTUPNÍCH DAT

Tabulka 4 -Vstupní data datového souboru DATA.CSV [3], [4], [21], [29], [30], [32]

Rok	Kraj	Stav obyvatelstva	Rozloha	Celkový přírůstek obyv.	Sňatky	Rozvody	Kriminalita celkem	Naděje na dožití - M	...
1998	1	10289621	7 8859.88	-9504	55027	32363	425930	71.1	...
1998	2	1193270	495.87	-7185	6512	4418	114070	72.3	...
1998	3	1108465	11014.38	2501	5808	3663	52440	70.6	...
1998	4	626634	10055.53	64	3467	1687	19577	70.8	...
1998	5	552553	7560.83	-716	3007	1622	16968	71.0	...
1998	6	304831	3314.39	-88	1807	1143	12625	70.0	...
1998	7	826852	5335.11	982	4525	3571	36155	68.8	...
1998	8	429080	3163.07	143	2467	1624	17419	70.3	...
1998	9	552481	4758.09	-371	3071	1647	14687	71.5	...
1998	10	509353	4518.43	-282	2727	1354	12348	71.4	...
1998	11	522267	6924.78	-349	2707	987	10210	71.4	...
1998	12	1138174	7062.25	-551	5741	3130	37602	71.6	...
1998	13	642961	5139.14	-985	3416	1871	19961	70.4	...
1998	14	598789	3964.04	-642	3098	1364	13474	70.7	...
1998	15	1283911	5553.97	-2025	6674	4282	48394	69.4	...
1999	1	10278098	78864.10	-11523	53523	23657	426626	71.4	...
1999	2	1186855	495.84	-6415	6441	3041	120166	72.9	...
1999	3	1111354	11014.32	2889	5849	2667	52191	70.8	...
1999	4	626112	10055.58	-522	3340	1641	19560	71.2	...
1999	5	551870	7560.82	-683	3006	1289	18446	71.6	...
1999	6	304823	3314.41	-8	1775	898	11713	70.7	...
1999	7	827151	5335.14	299	4361	1883	35078	69.2	...
1999	8	429012	3163.09	-68	2420	1050	16937	70.7	...
1999	9	551651	4758.11	-830	2939	1176	14306	71.9	...
1999	10	508744	4518.47	-609	2524	1180	12704	71.6	...
1999	11	521472	6925.20	-795	2639	947	9952	71.9	...
1999	12	1137289	7065.55	-885	5792	2317	37056	71.9	...
1999	13	642016	5139.27	-945	3118	1304	19091	71.0	...
1999	14	598339	3964.11	-450	2996	1113	13068	71.1	...
1999	15	1281410	5554.19	-2501	6323	3151	46358	69.9	...
2000	1	10266546	78865.25	-11552	55321	29704	391310	71.7	...
2000	2	1181126	495.89	-5729	6973	3842	106974	73.3	...
2000	3	1115038	11014.61	3684	6038	3303	45463	70.9	...
2000	4	625874	10056.34	-238	3423	1718	18641	72.0	...
2000	5	551281	7560.89	-589	3129	1672	17265	71.8	...
2000	6	304400	3314.40	-423	1847	1161	10675	70.8	...
2000	7	827013	5334.91	-138	4499	2717	32114	69.8	...
2000	8	429121	3163.12	109	2458	1307	16033	70.9	...
...

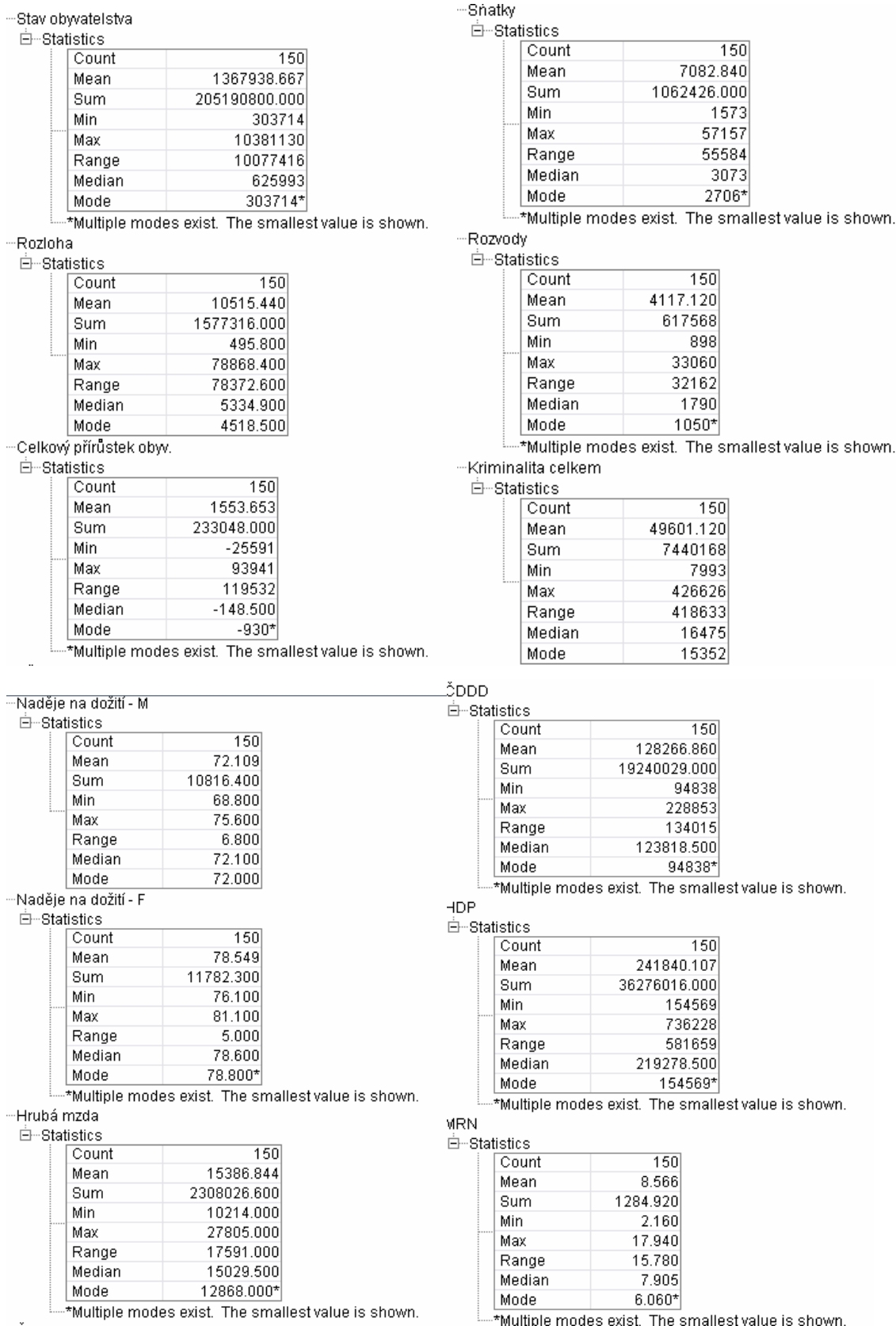
PŘÍLOHA č. 2: TYPY DAT

Obrázek 30 - Typy dat [Zdroj: vlastní]

Field	Type	Values	Missing	Check	Direction
Rok	Ordered Set	1998,1999...		None	In
Kraj	Set	"1","10","1...		None	In
Stav obyvatel...	Range	[303714,1...		None	In
Rozloha	Range	[495.8,788...		None	In
Celkový přírů...	Range	[-25591,93...		None	In
Sňatky	Range	[1573,571...		None	In
Rozvody	Range	[898,33060]		None	In
Kriminalita c...	Range	[7993,426...		None	In
Naděje na d...	Range	[68.8,75.6]		None	In
Naděje na d...	Range	[76.1,81.1]		None	In
Hrubá mzda	Range	[10214.0,2...		None	In
ČDDD	Range	[94838,22...		None	In
HDP	Range	[154569,7...		None	In
MRN	Range	[2.16,17.94]		None	In
Živě narození	Range	[2777,114...		None	In
Mrtvě narození	Range	[6,315]		None	In
Zemřelí	Range	[2851,111...		None	In
Nonotvary	Range	[819,29364]		None	In
Nemoci OS	Range	[1240,603...		None	In
Nemoci DS	Range	[104,6040]		None	In
Nemoci TS	Range	[115,4823]		None	In
Vnější příčiny	Range	[169,7295]		None	In
Počet nemoc...	Range	[5,225]		None	In
Počet lékařů	Range	[1021,436...		None	In
PIO ŽP	Range	[227951,4...		None	In
Tuhé látky	Range	[535.7,756...		None	In
SO2	Range	[1422.3,43...		None	In
NOx	Range	[1751.6,16...		None	In
CO	Range	[2117.8,39...		None	In

PŘÍLOHA č. 3: STATISTICKÁ ANALÝZA

Obrázek 31 - Statistická analýza vstupních dat [Zdroj: vlastní]



---Živě narození

▢ Statistics

Count	150
Mean	12912.533
Sum	1936880.000
Min	2777
Max	114632
Range	111855
Median	5724.500
Mode	4466*

*Multiple modes exist. The smallest value is shown.

---Nonotvary

▢ Statistics

Count	150
Mean	3800.867
Sum	570130.000
Min	819
Max	29364
Range	28545
Median	1737.500
Mode	824*

*Multiple modes exist. The smallest value is shown.

---Mrtvě narození

▢ Statistics

Count	150
Mean	37.573
Sum	5636.000
Min	6
Max	315
Range	309
Median	18
Mode	16

---Nemoci OS

▢ Statistics

Count	150
Mean	7556.227
Sum	1133434
Min	1240
Max	60397
Range	59157
Median	3359
Mode	3089

---Zemřelí

▢ Statistics

Count	150
Mean	14396.987
Sum	2159548.000
Min	2851
Max	111288
Range	108437
Median	6363.500
Mode	2851*

*Multiple modes exist. The smallest value is shown.

---Nemoci DS

▢ Statistics

Count	150
Mean	674.067
Sum	101110.000
Min	104
Max	6040
Range	5936
Median	318.500
Mode	295*

*Multiple modes exist. The smallest value is shown.

---Nemoci TS

▢ Statistics

Count	150
Mean	599.227
Sum	89884.000
Min	115
Max	4823
Range	4708
Median	283.500
Mode	183*

*Multiple modes exist. The smallest value is shown.

---Počet lékařů

▢ Statistics

Count	150
Mean	5283.953
Sum	792593.000
Min	1021
Max	43676
Range	42655
Median	2309.500
Mode	2541

---Vnější příčiny

▢ Statistics

Count	150
Mean	898.040
Sum	134706
Min	169
Max	7295
Range	7126
Median	432
Mode	305*

*Multiple modes exist. The smallest value is shown.

---PIO ŽP

▢ Statistics

Count	150
Mean	3198817.147
Sum	479822572.000
Min	227951
Max	44435229
Range	44207278
Median	975978
Mode	227951*

*Multiple modes exist. The smallest value is shown.

---Počet nemocnic

▢ Statistics

Count	150
Mean	26.707
Sum	4006.000
Min	5
Max	225
Range	220
Median	11
Mode	10

---Tuhé látky

▢ Statistics

Count	150
Mean	5681.078
Sum	852161.700
Min	535.700
Max	75699.600
Range	75163.900
Median	2394.850
Mode	4668.600

---SO2

▢ Statistics

Count	150
Mean	32610.287
Sum	4891543.000
Min	1422.300
Max	432038.200
Range	430615.900
Median	10897.950
Mode	1422.300*

*Multiple modes exist. The smallest value is shown.

---NOx

▢ Statistics

Count	150
Mean	20972.310
Sum	3145846.500
Min	1751.600
Max	164223.100
Range	162471.500
Median	4786.650
Mode	1751.600*

*Multiple modes exist. The smallest value is shown.

---CO

▢ Statistics

Count	150
Mean	36469.499
Sum	5470424.900
Min	2117.800
Max	397930.200
Range	395812.400
Median	8938.800
Mode	2117.800*

*Multiple modes exist. The smallest value is shown.

PŘÍLOHA č. 4: DATA AUDIT

Obrázek 32 - Kvalitativní analýza vstupních dat [Zdroj: vlastní]

Field	Graph	Type	Min	Max	Valid
Rok		Ordered Set	1998	2007	150
Kraj		Set	--	--	150
Stav obyvatelstva		Range	303714	10381130	150
Rozloha		Range	495.800	78868.400	150
Celkový přírůstek obyv.		Range	-25591	93941	150
Sňatky		Range	1573	57157	150
Rozvody		Range	898	33060	150
Kriminalita celkem		Range	7993	426626	150
Naděje na dožití - M		Range	68.800	75.600	150
Naděje na dožití - F		Range	76.100	81.100	150
Hrubá mzda		Range	10214.000	27805.000	150
ČDDD		Range	94838	228853	150
HDP		Range	154569	736228	150
MRN		Range	2.160	17.940	150
Živě narození		Range	2777	114632	150
Mrtvě narození		Range	6	315	150
Zemřelí		Range	2851	111288	150
Nonotvary		Range	819	29364	150
Nemoci OS		Range	1240	60397	150
Nemoci DS		Range	104	6040	150
Nemoci TS		Range	115	4823	150
Vnější příčiny		Range	169	7295	150

Field	Graph	Type	Min	Max	Valid
✚ Počet nemocnic		✚ Range	5	225	150
✚ Počet lékařů		✚ Range	1021	43676	150
✚ PIO ŽP		✚ Range	227951	44435229	150
✚ Tuhé látky		✚ Range	535.700	75699.600	150
✚ SO2		✚ Range	1422.300	432038.200	150
✚ NOx		✚ Range	1751.600	164223.100	150
✚ CO		✚ Range	2117.800	397930.200	150

PŘÍLOHA č. 5: VSTUPNÍ DATA DATOVÉHO SOUBORU „PRŮMĚRY“

Tabulka 5 - Vstupní data datového souboru PRŮMĚRY.CSV [Zdroj: vlastní]

Kraj	ŽN	MN	Z	N	NOS	NDS	NTS	VP	PL
1	96844.0	281.8	107977.4	28506.5	56671.7	5055.5	668249.4	6735.3	39630.0
2	10576.3	26.2	13078.1	3621.1	6753.4	631.9	521.3	783.2	7704.8
3	11056.2	29.1	12617.7	3314.3	6694.8	580.8	467.7	773.0	3381.6
4	5926.3	18.1	6371.3	1780.3	3193.2	315.9	232.5	405.9	2215.5
5	5122.5	16.0	5964.9	1693.3	3014.5	279.3	222.8	333.8	2283.2
6	2991.4	11.4	2975.6	859.4	1387.3	166.6	136.0	201.1	1083.7
7	8410.0	30.4	8957.5	2452.9	4584.4	353.4	400.4	577.6	2655.7
8	4231.4	12.7	4366.6	1208.2	2223.0	166.9	186.2	297.1	1409.9
9	5212.5	14.4	5786.7	1461.4	3122.7	305.0	196.6	358.0	2141.1
10	4855.8	13.9	5212.6	1282.4	2791.1	301.4	194.6	331.5	1635.7
11	4882.0	13.2	5198.8	1325.4	2943.0	215.6	165.5	289.2	1629.7
12	10205.7	27.0	11879.2	3027.9	6573.0	485.0	483.8	730.2	4671.7
13	5943.0	15.3	6551.1	1713.8	3383.0	296.2	309.7	438.9	2476.7
14	5374.7	13.4	6080.8	1425.2	3362.8	270.4	297.3	429.1	1891.0
15	11839.7	40.4	12947.0	3352.2	6677.6	668.8	675.0	796.5	4398.4

PŘÍLOHA č. 6: DATOVÝ SLOVNÍK DATOVÉHO SOUBORU „PRŮMĚRY“














Tabulka 6 - Datový slovník – PRŮMĚRY.CSV [Zdroj: vlastní]

č.	Ukazatel	Typ proměnné	Type	Rozsah	Měrná jedn.	Definice ukazatele
1	Rok	kategorizovaná	set	<1998;2007>	rok	označení období
2	Kraj	kategorizovaná	set	<1;15>	označení	označení oblasti (1-ČR;2-Hl.město Praha;3-Středočeský kraj;4-Jihočeský kraj;5-Plzeňský kraj;6-Karlovarský kraj;7-Ústecký kraj;8-Liberecký kraj;9-Královéhradecký kraj;10-Pardubický kraj;11-Vysočina;12-Jihomoravský kraj;13-Olomoucký kraj;14-Zlínský kraj;15-Moravskoslezský kraj)
3	ŽN	spojitá	range	<2777;114632>	osoby	počet hlášení o živě narozených
4	MN	spojitá	range	<6;315>	osoby	počet hlášení o mrtvě narozených
5	Z	spojitá	range	<2851;111288>	osoby	celkový počet zemřelých
6	N	spojitá	range	<819;29364>	osoby	počet zemřelých na novotvary
7	NOS	spojitá	range	<1240;60397>	osoby	počet zemřelých na nemoci oběhové soustavy
8	NDS	spojitá	range	<104;6040>	osoby	počet zemřelých na nemoci dýchací soustavy
9	NTS	spojitá	range	<115;4823>	osoby	počet zemřelých na nemoci trávicí soustavy
10	VP	spojitá	range	<169;7295>	osoby	počet zemřelých na vnější příčiny
11	PL	spojitá	range	<1021;43676>	osoby	aktuální počet lékařů

PŘÍLOHA č. 7: ROZDĚLENÍ DO SHLUKŮ PODLE ZDRAVOTNÍ ÚROVNĚ

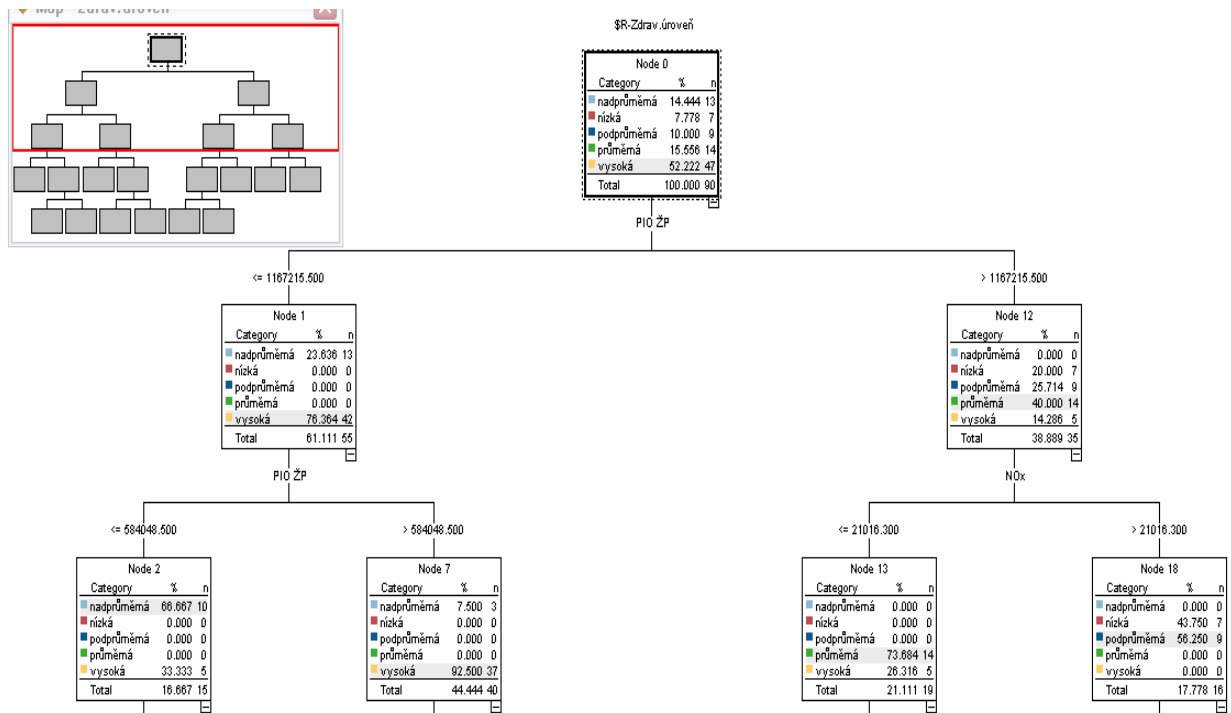
Obrázek 33 - Rozdělení do shluků - zdravotní úroveň [Zdroj: vlastní]

K-Means
KMeans

	cluster-1 Count: 70 (53%)	cluster-4 Count: 20 (15%)	cluster-5 Count: 20 (15%)	cluster-2 Count: 10 (7%)	cluster-3 Count: 10 (7%)	Importance 					
Kraje	Hl. měs...	0 0.00%	Hl. měs...	0 0.00%	Hl. měs...	0 0.00%	Hl. měs...	0 0.00%	Hl. měs...	0 0.00%	Unimportant 0.00 
	Jihomor...	0 0.00%	Jihomor...	10 50.00%	Jihomor...	0 0.00%	Jihomor...	0 0.00%	Jihomor...	0 0.00%	
	Jihočes...	10 14.29%	Jihočes...	0 0.00%	Jihočes...	0 0.00%	Jihočes...	0 0.00%	Jihočes...	0 0.00%	
	Karlova...	0 0.00%	Karlova...	0 0.00%	Karlova...	10 50.00%	Karlova...	0 0.00%	Karlova...	0 0.00%	
	Králové...	10 14.29%	Králové...	0 0.00%	Králové...	0 0.00%	Králové...	0 0.00%	Králové...	0 0.00%	
	Liberec...	0 0.00%	Liberec...	0 0.00%	Liberec...	10 50.00%	Liberec...	0 0.00%	Liberec...	0 0.00%	
	Moravsk...	0 0.00%	Moravsk...	0 0.00%	Moravsk...	0 0.00%	Moravsk...	10 100.00%	Moravsk...	0 0.00%	
	Olomouc...	10 14.29%	Olomouc...	0 0.00%	Olomouc...	0 0.00%	Olomouc...	0 0.00%	Olomouc...	0 0.00%	
	Pardubi...	10 14.29%	Pardubi...	0 0.00%	Pardubi...	0 0.00%	Pardubi...	0 0.00%	Pardubi...	0 0.00%	
	Plzeňsk...	10 14.29%	Plzeňsk...	0 0.00%	Plzeňsk...	0 0.00%	Plzeňsk...	0 0.00%	Plzeňsk...	0 0.00%	
	Středoč...	0 0.00%	Středoč...	10 50.00%	Středoč...	0 0.00%	Středoč...	0 0.00%	Středoč...	0 0.00%	
	Vysočina	10 14.29%	Vysočina	0 0.00%	Vysočina	0 0.00%	Vysočina	0 0.00%	Vysočina	0 0.00%	
	Zlínský...	10 14.29%	Zlínský...	0 0.00%	Zlínský...	0 0.00%	Zlínský...	0 0.00%	Zlínský...	0 0.00%	
	Ústecký...	0 0.00%	Ústecký...	0 0.00%	Ústecký...	0 0.00%	Ústecký...	0 0.00%	Ústecký...	10 100.00%	
	Česká r...	0 0.00%	Česká r...	0 0.00%	Česká r...	0 0.00%	Česká r...	0 0.00%	Česká r...	0 0.00%	
MN na 10000 obyv	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	Important 1.00 				
N na 10000 obyv	0.15 (0.02)	0.32 (0.02)	0.10 (0.02)	0.34 (0.00)	0.25 (0.00)	0.25 (0.00)	Important 1.00 				
NDS na 10000 obyv	0.03 (0.00)	0.05 (0.00)	0.02 (0.00)	0.07 (0.00)	0.04 (0.00)	0.04 (0.00)	Important 1.00 				
NOS na 10000 obyv	0.31 (0.02)	0.86 (0.01)	0.18 (0.04)	0.87 (0.00)	0.46 (0.00)	0.46 (0.00)	Important 1.00 				
NTS na 10000 obyv	0.02 (0.01)	0.05 (0.00)	0.02 (0.00)	0.07 (0.00)	0.04 (0.00)	0.04 (0.00)	Important 1.00 				
Naděje na dožití - F	78.85 (0.72)	78.74 (0.83)	77.88 (0.79)	78.07 (0.63)	78.86 (0.61)	78.86 (0.61)	Important 1.00 				
Naděje na dožití - M	72.47 (0.88)	72.22 (0.85)	71.43 (0.84)	70.73 (0.90)	70.12 (0.83)	70.12 (0.83)	Important 1.00 				
PL na 10000 obyv	0.20 (0.03)	0.41 (0.07)	0.12 (0.02)	0.44 (0.00)	0.27 (0.00)	0.27 (0.00)	Important 1.00 				
VP na 10000 obyv	0.04 (0.01)	0.07 (0.00)	0.02 (0.00)	0.08 (0.00)	0.06 (0.00)	0.06 (0.00)	Important 1.00 				
Zna 10000 obyv	0.59 (0.05)	1.22 (0.04)	0.37 (0.07)	1.29 (0.00)	0.90 (0.00)	0.90 (0.00)	Important 1.00 				
ŽN na 10000 obyv	0.53 (0.04)	1.07 (0.03)	0.38 (0.06)	1.18 (0.00)	0.84 (0.00)	0.84 (0.00)	Important 1.00 				

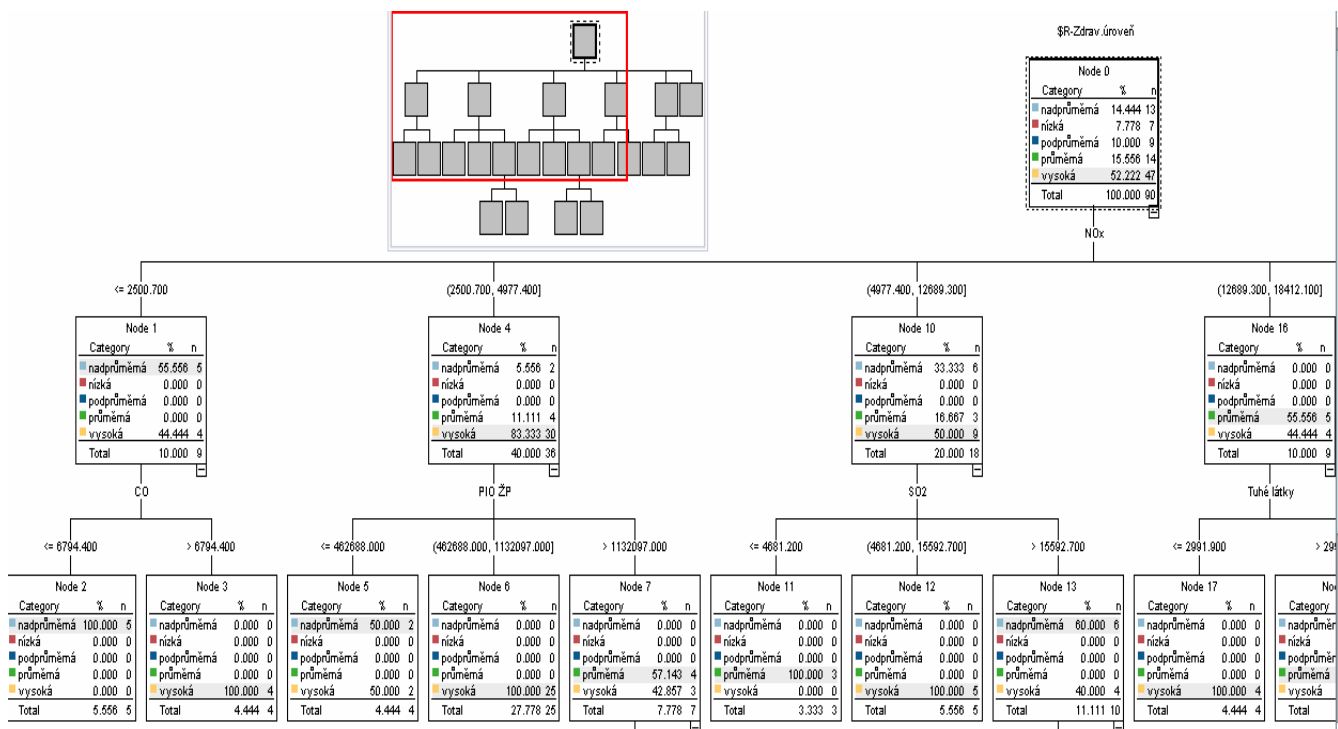
PŘÍLOHA č. 8: ROZHODOVACÍ STROM C&RT

Obrázek 34 - Detail RS C&RT [Zdroj: vlastní]



PŘÍLOHA č. 9: ROZHODOVACÍ STROM CHAID










Obrázek 35 - Detail RS CHAID [Zdroj: vlastní]



PŘÍLOHA č. 10: ROZDĚLENÍ DO SHLUKŮ PODLE EKONOMICKÉ ÚROVNĚ

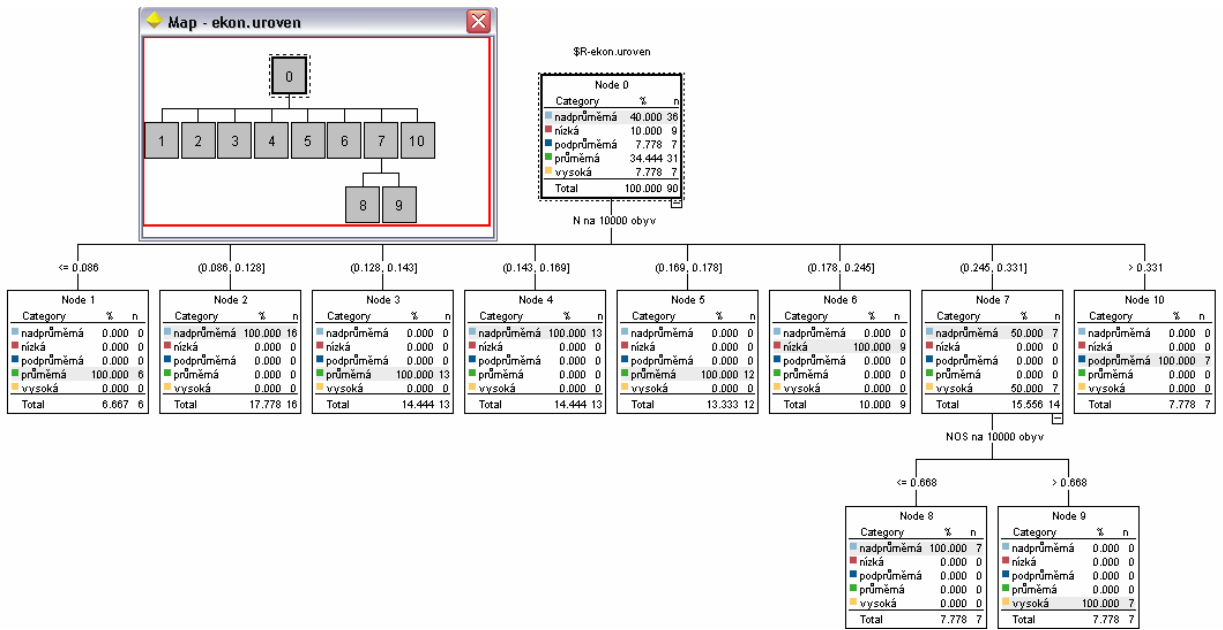
Obrázek 36 - Rozdělení do shluků - ekonomická úroveň [Zdroj: vlastní]

K-Means
KMeans

	cluster-1 Count: 50 (38%)	cluster-4 Count: 50 (38%)	cluster-2 Count: 10 (7%)	cluster-3 Count: 10 (7%)	cluster-5 Count: 10 (7%)	Importance  >=0,95  >=0,90  <0,90  Unknown
HDP	228828.54 (39340.62)	210840.32 (38652.73)	239813.94 (47476.85)	208719.59 (37104.58)	207421.50 (46372.47)	Marginal  0.92
Hrubá mzda	14835.63 (2827.40)	14382.21 (2659.95)	16259.18 (3406.46)	15066.58 (2831.88)	15554.06 (2910.85)	Unimportant  0.86
Kraje	H. měs.. 0 0.00%	H. měs.. 0 0.00%	H. měs.. 0 0.00%	H. měs.. 0 0.00%	H. měs.. 0 0.00%	Unimportant  0.00
	Jihomor.. 10 20.00%	Jihomor.. 0 0.00%	Jihomor.. 0 0.00%	Jihomor.. 0 0.00%	Jihomor.. 0 0.00%	
	Jihočes.. 0 0.00%	Jihočes.. 10 20.00%	Jihočes.. 0 0.00%	Jihočes.. 0 0.00%	Jihočes.. 0 0.00%	
	Karlova.. 0 0.00%	Karlova.. 10 20.00%	Karlova.. 0 0.00%	Karlova.. 0 0.00%	Karlova.. 0 0.00%	
	Králové.. 10 20.00%	Králové.. 0 0.00%	Králové.. 0 0.00%	Králové.. 0 0.00%	Králové.. 0 0.00%	
	Liberec.. 10 20.00%	Liberec.. 0 0.00%	Liberec.. 0 0.00%	Liberec.. 0 0.00%	Liberec.. 0 0.00%	
	Moravsk.. 0 0.00%	Moravsk.. 0 0.00%	Moravsk.. 0 0.00%	Moravsk.. 0 0.00%	Moravsk.. 10 100.00%	
	Olomouc.. 0 0.00%	Olomouc.. 10 20.00%	Olomouc.. 0 0.00%	Olomouc.. 0 0.00%	Olomouc.. 0 0.00%	
	Pardubi.. 10 20.00%	Pardubi.. 0 0.00%	Pardubi.. 0 0.00%	Pardubi.. 0 0.00%	Pardubi.. 0 0.00%	
	Plzeňsk.. 10 20.00%	Plzeňsk.. 0 0.00%	Plzeňsk.. 0 0.00%	Plzeňsk.. 0 0.00%	Plzeňsk.. 0 0.00%	
	Středoč.. 0 0.00%	Středoč.. 0 0.00%	Středoč.. 10 100.00%	Středoč.. 0 0.00%	Středoč.. 0 0.00%	
	Vysočina 0 0.00%	Vysočina 10 20.00%	Vysočina 0 0.00%	Vysočina 0 0.00%	Vysočina 0 0.00%	
	Zlínský.. 0 0.00%	Zlínský.. 10 20.00%	Zlínský.. 0 0.00%	Zlínský.. 0 0.00%	Zlínský.. 0 0.00%	
	Ústecký.. 0 0.00%	Ústecký.. 0 0.00%	Ústecký.. 0 0.00%	Ústecký.. 10 100.00%	Ústecký.. 0 0.00%	
Česká r.. 0 0.00%	Česká r.. 0 0.00%	Česká r.. 0 0.00%	Česká r.. 0 0.00%	Česká r.. 0 0.00%		
MRN	7.67 (1.58)	8.51 (2.04)	6.44 (1.01)	15.16 (2.03)	14.14 (2.24)	Important  1.00
ČDDD	126082.93 (18208.50)	122434.77 (18717.16)	138883.50 (23325.49)	116820.48 (16466.89)	118094.96 (17943.91)	Marginal  0.94

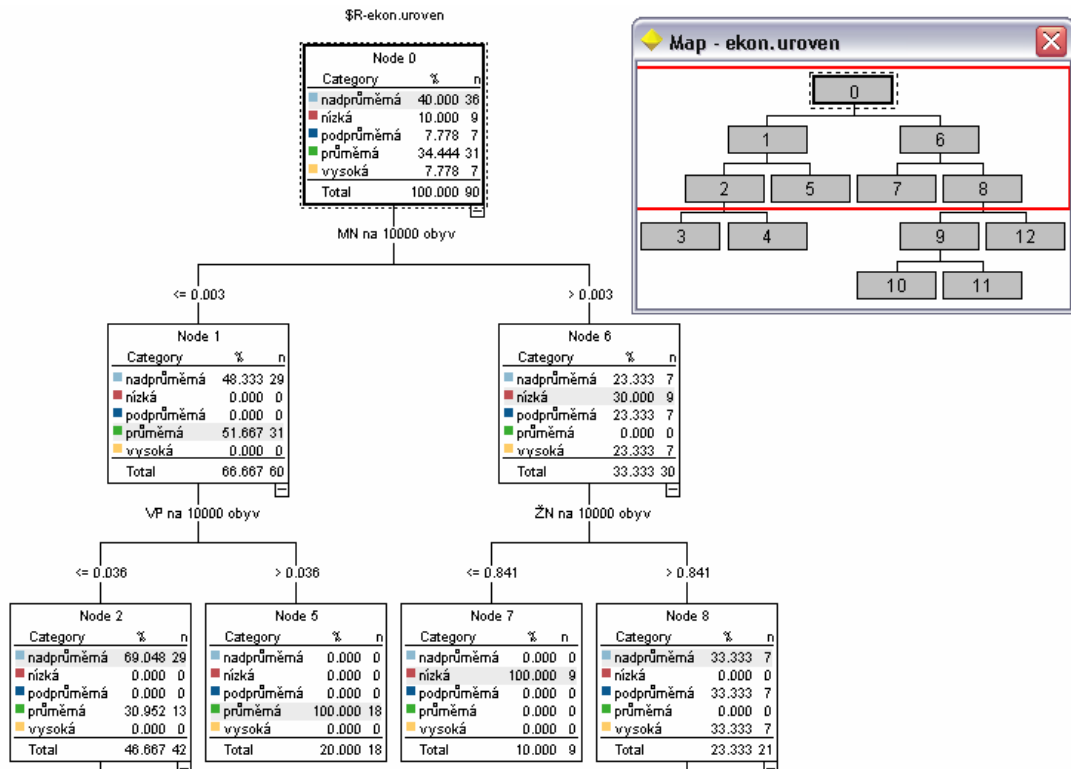
PŘÍLOHA č. 11: ROZHODOVACÍ STROM CHAID

Obrázek 37 - Detail RS CHAID [Zdroj: vlastní]



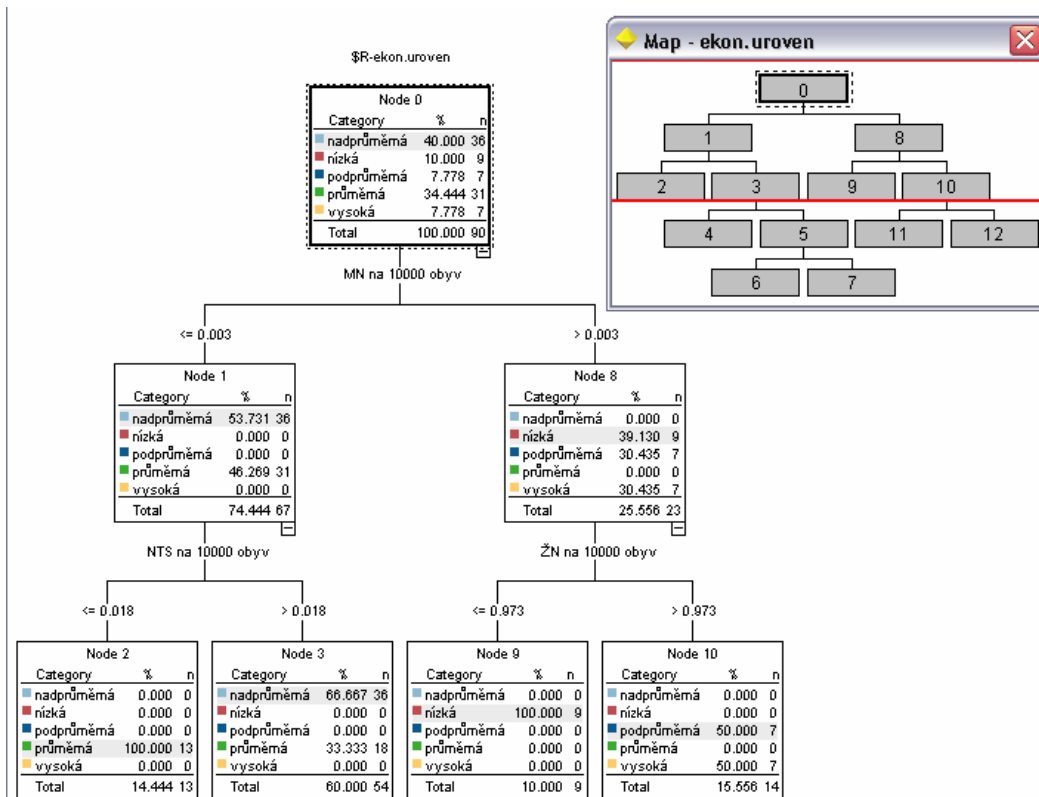
PŘÍLOHA č. 12: ROZHODOVACÍ STROM QUEST

Obrázek 38 - Detail RS QUEST [Zdroj: vlastní]












PŘÍLOHA č. 13: ROZHODOVACÍ STROM C&RT

Obrázek 39 - Detail RS C&RT [Zdroj: vlastní]



PŘÍLOHA č. 14: ROZDĚLENÍ DO SHLUKŮ PODLE VYBRANÝCH UKAZATELŮ

Obrázek 40 - Rozdělení do shluků - vybrané ukazatele [Zdroj: vlastní]

		K-Means			K-Means			Importance			
		cluster-1			cluster-3			cluster-2			
		Count: 90 (89 %)			Count: 30 (23 %)			Count: 10 (7 %)			
								 >=0.95  >=0.90  <0.90  Unknown			
Celkový přírůstek obyv.		156.31 (1367.80)			-304.72 (2801.95)			8931.04 (7992.11)			Important  1.00
Kraje	H. měs...	0	0.00 %	H. měs...	0	0.00 %	H. měs...	0	0.00 %	Unimportant  0.00	
	Jihomor...	0	0.00 %	Jihomor...	10	33.33 %	Jihomor...	0	0.00 %		
	Jihočes...	10	11.11 %	Jihočes...	0	0.00 %	Jihočes...	0	0.00 %		
	Karlova...	10	11.11 %	Karlova...	0	0.00 %	Karlova...	0	0.00 %		
	Králové...	10	11.11 %	Králové...	0	0.00 %	Králové...	0	0.00 %		
	Liberec...	10	11.11 %	Liberec...	0	0.00 %	Liberec...	0	0.00 %		
	Moravsk...	0	0.00 %	Moravsk...	10	33.33 %	Moravsk...	0	0.00 %		
	Olomouc...	10	11.11 %	Olomouc...	0	0.00 %	Olomouc...	0	0.00 %		
	Pardubi...	10	11.11 %	Pardubi...	0	0.00 %	Pardubi...	0	0.00 %		
	Plzeňsk...	10	11.11 %	Plzeňsk...	0	0.00 %	Plzeňsk...	0	0.00 %		
	Středoč...	0	0.00 %	Středoč...	0	0.00 %	Středoč...	10	100.0...		
	Vysočina	10	11.11 %	Vysočina	0	0.00 %	Vysočina	0	0.00 %		
	Žitavský...	10	11.11 %	Žitavský...	0	0.00 %	Žitavský...	0	0.00 %		
Ústecký...	0	0.00 %	Ústecký...	10	33.33 %	Ústecký...	0	0.00 %			
Česká r...	0	0.00 %	Česká r...	0	0.00 %	Česká r...	0	0.00 %			
Kriminalita celkem	13479.90 (3097.78)			35524.36 (4317.51)			42806.31 (5760.99)			Important  1.00	
Rozvody	1495.20 (278.53)			3357.37 (611.29)			3535.00 (350.76)			Important  1.00	
Sňatky	2676.23 (453.63)			5467.00 (800.45)			5966.90 (291.70)			Important  1.00	