

Univerzita Pardubice
Fakulta ekonomicko-správní

**Návrh klasifikátoru parametrů udržitelného rozvoje
pomocí Support Vector Machine**

Bc. Petra Špírková

Diplomová práce

2010

Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky
Akademický rok: 2009/2010

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Petra ŠPÍRKOVÁ**

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

Studijní obor: **Informatika ve veřejné správě**

Název tématu: **Návrh klasifikátoru parametrů udržitelného rozvoje pomocí Support Vector Machine**

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

Charakteristika udržitelného rozvoje.

Analyza vstupních dat (parametrů) pro následující klasifikaci.

Charakteristika SVM z hlediska klasifikace a aproximace.

Návrh modelu na modelování udržitelného rozvoje.

Verifikace navrženého modelu.

Analýza výsledků.

Rozsah grafických prací:

Rozsah pracovní zprávy:

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

Seznam odborné literatury:

1. OLEJ, V.: Modelovanie ekonomických procesov na báze výpočtovej inteligencie. [Vedecká monografia], Miloš Vognar - M&V, ISBN 80-903024-9-1, Hradec Králové, Česká republika, 2003, 160s.
2. KVASNIČKA, V. a kol.: Úvod do teórie neurónových sietí. Iris, Bratislava, 1997.
3. HAYKIN, S.: Neural Networks: A Comprehensive Foundation. 2nd edition, New Jersey, Prentice-Hall, Inc., 1999, 842s.
4. FIALOVÁ, H. a kol.: Vybrané oblasti udržitelného rozvoje v Pardubickém kraji. Český statistický úřad, Oddelení regionálních analýz a informačních služeb Pardubice, Pardubice, 2007.



Vedoucí diplomové práce:

prof. Ing. Vladimír Olej, CSc.
Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce: **5. října 2009**

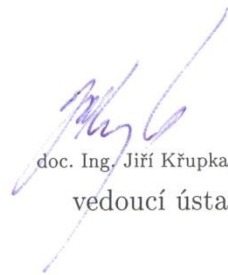
Termín odevzdání diplomové práce: **30. dubna 2010**



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 _____

Petra Špírková

PODĚKOVÁNÍ

Touto cestou bych ráda poděkovala především vedoucímu diplomové práce, panu prof. Ing. Vladimíru Olejovi, CSc. za odbornou pomoc, cenné rady a připomínky k obsahové i formální stránce této diplomové práce, za čas, který mi věnoval, za vedení a podporu vedení při vypracování.

ANOTACE

Diplomová práce je zaměřena na návrh klasifikátoru parametrů udržitelného rozvoje pomocí Support Vector Machine. Řešená problematika se vztahuje na vybranou oblast Pardubického kraje. V první části práce je popsána problematika udržitelného rozvoje a metoda Support Vector Machine. Další část práce je věnována popisu vstupních dat použitých k modelování, předzpracování dat a následnému návrhu vhodného klasifikátoru pomocí metody Support Vector Machine. V závěru jsou analyzovány a popsány výsledky navrženého modelu a aproximace.

KLÍČOVÁ SLOVA

udržitelný rozvoj; indikátory; Support Vector Machine; programový prostředek Weka; klasifikace; aproximace

TITLE

Classifier of parameters sustainable development design by Support Vector Machine

ANNOTATION

This thesis deals with classifier of parameters sustainable development design by Support Vector Machine. Dealing with the issue applies to the selected area of the Pardubice region. In the first part of the work the problem of sustainable development and Support Vector Machine method has been reported. Next part is dedicated to data description, which were used for modelling, data processing and proposal of optimal classifier by Support Vector Machine method. Finally, the designed model and approximation results have been analyzed and described.

KEYWORDS

sustainable development; parametres; Support Vector Machine; Weka programing; classification; approximation

OBSAH

ÚVOD.....	8
1 UDRŽITELNÝ ROZVOJ	9
1.1 VYMEZENÍ UDRŽITELNÉHO ROZVOJE	9
1.2 STRATEGIE UDRŽITELNÉHO ROZVOJE	10
1.2.1 <i>Strategie udržitelného rozvoje Evropské unie</i>	10
1.2.2 <i>Strategie udržitelného rozvoje České republiky</i>	12
1.3 PILÍŘE UDRŽITELNÉHO ROZVOJE	13
1.4 PRINCIPY UDRŽITELNÉHO ROZVOJE	14
1.5 INDIKÁTORY UDRŽITELNÉHO ROZVOJE	15
1.6 DÍLČÍ ZÁVĚR.....	17
2 SUPPORT VECTOR MACHINE	18
2.1 ÚVOD DO PROBLEMATIKY SUPPORT VECTOR MACHINE.....	18
2.2 OPTIMÁLNÍ NADROVINA PRO LINEÁRNĚ SEPAROVANÁ DATA	19
2.3 OPTIMÁLNÍ NADROVINA PRO LINEÁRNĚ NESEPAROVANÁ DATA.....	21
2.4 DÍLČÍ ZÁVĚR.....	23
3 ANALÝZA VSTUPNÍCH DAT	24
3.1 PARDUBICKÝ KRAJ	24
3.2 INDIKÁTORY UDRŽITELNÉHO ROZVOJE	25
3.2.1 <i>Ekonomické indikátory</i>	25
3.2.2 <i>Environmentální indikátory</i>	27
3.2.3 <i>Sociální indikátory</i>	29
3.3 DÍLČÍ ZÁVĚR.....	35
4 NÁVRH MODELU KLASIFIKÁTORU UDRŽITELNÉHO ROZVOJE	36
4.1 NÁVRH MODELU PRO KLASIFIKACI OBCÍ.....	36
4.2 CHARAKTERISTIKA VSTUPNÍCH DAT	37
4.3 PŘEDZPRACOVÁNÍ DAT	39
4.4 ROZDĚLENÍ DAT NA TESTOVACÍ A TRÉNOVACÍ MNOŽINU	39
4.5 ALGORITMUS SEQUENTIAL MINIMAL OPTIMIZATION.....	41
4.6 NAVRŽENÝ MODEL V PROGRAMOVÉM PROSTŘEDÍ WEKA.....	41
4.7 DÍLČÍ ZÁVĚR.....	42

5	ANALÝZA NAVRŽENÉHO MODELU	43
5.1	PROGRAMOVÉ PROSTŘEDÍ WEKA	43
5.2	IMPORT DAT.....	43
5.3	NÁVRH MODELU V PROGRAMOVÉM PROSTŘEDÍ WEKA KNOWLEDGEFLOW.....	44
5.4	NÁVRH PARAMETRŮ MODELU PRO EKONOMICKÁ DATA.....	45
5.4.1	<i>Optimální struktura navrženého modelu pro ekonomická data.....</i>	<i>46</i>
5.4.2	<i>Analýza klasifikátoru zvoleného modelu pro ekonomická data.....</i>	<i>47</i>
5.5	NÁVRH PARAMETRŮ MODELU PRO ENVIRONMENTÁLNÍ DATA	49
5.5.1	<i>Optimální struktura navrženého modelu pro environmentální data</i>	<i>50</i>
5.5.2	<i>Analýza klasifikátoru zvoleného modelu pro environmentální data.....</i>	<i>50</i>
5.6	NÁVRH PARAMETRŮ MODELU PRO SOCIÁLNÍ DATA	52
5.6.1	<i>Optimální struktura navrženého modelu pro sociální data</i>	<i>53</i>
5.6.2	<i>Analýza klasifikátoru zvoleného modelu pro sociální data</i>	<i>54</i>
5.7	DÍLČÍ ZÁVĚR.....	56
6	APROXIMACE NAVRŽENÝCH MODELŮ	57
6.1	KORELACE.....	57
6.2	ANALÝZA APROXIMACE	58
6.3	APROXIMACE EKONOMICKÉHO MODELU.....	59
6.4	APROXIMACE ENVIRONMENTÁLNÍHO MODELU	61
6.5	APROXIMACE SOCIÁLNÍHO MODELU	67
6.6	ZHODNOCENÍ VÝSLEDKŮ NAVRŽENÝCH MODELŮ	74
6.7	DÍLČÍ ZÁVĚR.....	75
	ZÁVĚR	76
	SEZNAM POUŽITÉ LITERATURY.....	77
	SEZNAM ZKRATEK	80
	SEZNAM OBRÁZKŮ	81
	SEZNAM GRAFŮ	82
	SEZNAM TABULEK.....	83
	SEZNAM PŘÍLOH.....	84

ÚVOD

Problematika udržitelného rozvoje se v posledních letech stává stále více diskutovaným tématem. Významným faktorem, který tuto problematiku ovlivňuje, je neustálá změna životního prostředí a snaha zachování množství přírodních zdrojů pro budoucí generace.

Cílem diplomové práce bude navržení klasifikátoru parametrů udržitelného rozvoje na základě využití metody Support Vector Machine a ověření správnosti navržení parametrů expertem.

V první části diplomové práce bude přiblížena problematika udržitelného rozvoje a popsána metoda Support Vector Machine, která byla pro návrh klasifikátoru vybrána. V rámci problematiky udržitelného rozvoje bude vysvětlen pojem udržitelného rozvoje a popsány významné strategie vztahující se v území České republiky. Dále bude věnována pozornost principům a základním parametrům udržitelného rozvoje. V závěru udržitelného rozvoje bude práce zaměřena na definování základních pilířů a indikátorů udržitelného rozvoje. V závěru této části bude pozornost zaměřena na teoretický popis zvolené metody Support Vector Machine se zaměřením na úlohu klasifikace.

Druhá část této práce bude zaměřena na navržení modelu klasifikace pro parametry, které byly vybrány expertem, v programovém prostředí Weka. V důsledku velkého množství stanovených parametrů bude sestaven pro každou oblast dat vlastní model. Na základě navržených modelů bude provedena analýza výsledků, v rámci níž budou specifikovány důvody nastavených hodnot klasifikátoru.

V závěrečné části diplomové práce bude provedena aproximace navržených a následná analýza. Cílem aproximace je získat nejlepší přesnost klasifikace na základě navržení některých parametrů na vyřazení, pokud by jejich vyřazení přispělo ke zlepšení přesnosti. V závislosti na aproximaci budou získány informace, zda nastavení parametrů pro klasifikaci od experta bylo vhodně navrženo.

1 Udržitelný rozvoj

V posledních letech se pojem udržitelný rozvoj vyskytuje velmi často v mnoha národních i mezinárodních dokumentech. Udržitelný rozvoj neznamena pouze ochranu životního prostředí a efektivní využívání přírodních zdrojů, ale také zahrnuje aspekty ekonomické a sociální soudržnosti [2].

V této kapitole jsou dále blíže specifikovány základní pojmy z oblasti udržitelného rozvoje. Podstatou celé kapitoly je vymezení pojmu udržitelného rozvoje, popis významných strategií, přiblížení principů, pilířů a indikátory udržitelného rozvoje.

1.1 Vymezení udržitelného rozvoje

Existuje celá řada definic charakterizující udržitelný rozvoj. Na téma udržitelného rozvoje byla a je vedena celá řada diskuzí. Na jejich základě Valné shromáždění Organizace spojených národů (OSN) v roce 1983 ustanovilo Světovou komisi pro životní prostředí a rozvoj (WCED). Tato komise byla pověřena analyzováním vztahu mezi hospodářským vývojem a životním prostředím a navržením způsobu, jak mezi nimi překonat zdánlivě neřešitelné rozpory. Komise pracovala tři roky pod vedením norské ministerské předsedkyně Gro Garlem Brundtlandové a v roce 1987 vydala zprávu: „Naše společná budoucnost“.

Ve zprávě byl zveřejněn návrh, aby problémy životního prostředí byly v celosvětovém měřítku řešeny nikoliv zastavením, ale novým typem hospodářského rozvoje – „rozvojem trvale udržitelným“ [21]. Základní aspekt udržitelného rozvoje je vystižen v definici uvedené zprávy: „Jedná se o takový udržitelný rozvoj, který zajistí naplnění potřeb současné generace, aniž by ohrozil možnost splnění potřeb generací příštích“ [4].

Trvalá udržitelnost zahrnuje dlouhodobou a globální perspektivu a představuje integraci čtyř základních cílů, jichž je potřeba dosáhnout pokud možno zároveň. Mezi tyto cíle je zahrnován sociální pokrok, účinná ochrana životního prostředí, uvážlivé využívání přírodních zdrojů a udržení vysoké a stabilní úrovně ekonomického růstu [11].

Vedle dosahování rovnováhy mezi sociální, ekonomickou a environmentální oblastí našich životů je další důležitou součástí dosažení také rovnováhy mezi zeměmi, různými společenskými skupinami, dneškem a budoucností [2].

1.2 Strategie udržitelného rozvoje

Udržitelný rozvoj je představován jako komplexní soubor strategií, které umožňují pomocí ekonomických prostředků a technologií uspokojovat lidské potřeby, materiální, kulturní i duchovní, při plném respektování environmentálních limitů [2].

Každá strategie udržitelného rozvoje určuje priority udržitelného rozvoje území a přiřazuje k nim dlouhodobé cíle, kterých by mělo být dosaženo v určitém časovém horizontu. Cíle udržitelného rozvoje bývají často velmi obecné, ať už se jedná o globální, národní, regionální či místní úroveň. Ke konkrétnosti dochází až při stanovení opatření jak cílů dosáhnout. Pro hodnocení vývoje udržitelnosti a úspěšnosti jednotlivých opatření jsou dále stanoveny konkrétní indikátory [21].

V rámci této podkapitoly je dobré uvést některé významné strategie, které se úzce vztahují k České republice. Mezi tyto strategie patří Strategie udržitelného rozvoje Evropské unie a Strategie udržitelného rozvoje České republiky, které budou následně blíže specifikovány.

1.2.1 Strategie udržitelného rozvoje Evropské unie

Evropská unie (EU) vydala v roce 2006 Strategii udržitelného rozvoje EU, která navazuje na dlouhodobé cíle, které byly stanoveny v první strategii Evropské unie přijaté v roce 2001. Vyjadřuje reakci na přetrvávající neudržitelné trendy, zejména v případě změn klimatu a spotřeby energie. Rovněž se nepodařilo zvrátit některé negativní jevy, mezi které je možno zahrnout hrozby pro veřejné zdraví, důsledky chudoby a sociálního vyloučení či stárnutí populace. Z těchto důvodů základní cíle, které byly obsaženy v první strategii, platí i nadále [4].

Aby došlo ke změně uvedených negativních jevů, bylo vyzdviženo ve druhé strategii sedm klíčových oblastí [4]:

- **Změna klimatu a čistá energie**

Stanoveným cílem této oblasti je minimalizovat změnu klimatu, náklady, které vyvolává, a její negativní efekty na společnost a životní prostředí. Týká se především energetické a dopravní politiky a hlavními směry jsou zvyšování podílu energie vyrobené z obnovitelných zdrojů a podílu zelených paliv.

- **Udržitelná doprava**

Druhá oblast strategie je zaměřena na dopravní systémy. Ty by měly vyhovovat ekonomickým a sociálním potřebám společnosti a přitom minimálně zatěžovat společnost, ekonomiku a životní prostředí. Cílem je oddělit ekonomický růst od růstu dopravy, snížit nepříznivé dopady ekonomického růstu a dopravy na životní prostředí a omezit emise skleníkových plynů v dopravě na úroveň, která minimalizuje škody pro zdraví obyvatel.

- **Udržitelná spotřeba a výroba**

Tato oblast představuje podporu udržitelných spotřebních zvyklostí a výrobních metod. Důležitým cílem je zmírnit závislost ekonomického růstu na poškozování životního prostředí a zvyšovat podíl technologií šetrných k životnímu prostředí a podíl ekologických inovací. Do této oblasti je také zahrnuta podpora ekologického zemědělství a mezinárodního obchodu založeného na spravedlivém přístupu, kdy nedochází k diskriminaci výrobců.

- **Šetrné nakládání s přírodními zdroji**

Čtvrtá oblast strategie se zaměřuje na zlepšení postupů při nakládání s přírodními zdroji a na ochranu před jejich nadměrným využíváním. Tento cíl se týká především lovu ryb, vody, půdy, vzduchu a atmosféry.

- **Veřejné zdraví**

Smyslem oblasti Veřejné zdraví je podporovat veřejné zdravotnictví založené na rovných podmínkách pro všechny a na lepší ochraně před možnými zdravotními ohroženími. S tím souvisí také přiměřená prevence zdraví, zlepšení legislativy týkající se zdravotní nezávadnosti potravin, boj proti vzestupu výskytu chronických onemocnění a nemocí v důsledku špatného životního stylu, zvláště u sociálně slabých skupin obyvatelstva.

- **Sociální začlenění, demografie a migrace**

Základním nástrojem této oblasti je podpora solidarity mezi generacemi a v rámci generací, úsilí o vyšší kvalitu života občanů, která je základní podmínkou pro blahobyt jednotlivce. Další součástí je modernizace sociální ochrany vzhledem k demografickým změnám, zvyšování a snižování negativních dopadů globalizace na pracovníky a jejich rodiny.

- **Globální chudoba a výzvy udržitelného rozvoje**

Cílem poslední oblasti je aktivní podpora udržitelného rozvoje v celosvětovém měřítku a zajištění, aby vnitřní a vnější politika EU byla v souladu s globálním udržitelným rozvojem a s jejími mezinárodními závazky.

1.2.2 Strategie udržitelného rozvoje České republiky

Česká republika (ČR) jako člen Evropské unie (EU) a mezinárodního společenství demokraticky a ekonomicky vyspělých zemí uznává princip solidarity mezi lidmi a mezi státy navzájem a přijímá svůj díl odpovědnosti při řešení globálních problémů. Jedním z výrazů tohoto postoje je zahraniční rozvojová spolupráce (ZRS), která je integrální součástí české zahraniční politiky.

Mezi hlavní cíle české ZRS patří snížení chudoby, ekonomicko-průmyslový rozvoj, postupná integrace partnerských zemí do světové ekonomiky, rozvoj zemědělství, rozvoj a upevnění demokracie, lidských práv a řádné správy věcí veřejných, zavedení právních principů, řízení migrace a udržitelný rozvoj s důrazem na jeho environmentální a sociální složku [17].

Česká republika (ČR) zareagovala na světovou tendenci a v roce 2004 vydala Strategii udržitelného rozvoje. V aktualizované Strategii udržitelného rozvoje ČR z roku 2007 jsou vymezeny prioritní oblasti, hlavní cíle a nástroje na jejich dosažení. Základem strategie zůstává minimalizovat prostor pro vznik nerovnováhy mezi ekonomickým, environmentálním a sociálním pilířem, což je považováno za základ jejich optimálního vývoje. Dalším úkolem je upozorňovat na existující a potenciální překážky udržitelného rozvoje a navrhnout vhodná opatření, jak zmírnit případné negativní dopady.

Strategie se opírá o několik principů, z nichž nejdůležitější je úcta k lidskému životu a k přírodě, k civilizačním a kulturním hodnotám. Dalšími významnými principy jsou sociální soudržnost a solidarita, rovné příležitosti, pozitivní ekonomické stimuly a subsidiarita¹. Ve strategii jsou tyto principy uplatňovány pro vybrané úseky společenského vývoje[4].

¹ Funkce má být přidělena vždy té úrovni veřejné správy, která má pro její naplňování co nejlepší informační zázemí.

K základním strategickým cílům je řazen stabilní ekonomický vývoj, který nadměrně nezatěžuje životní prostředí a přiměřeně reaguje na důsledky globalizačních procesů, rozvoj ekonomiky založené na znalostech², existence účinného rámce pro zvyšování konkurenční schopnosti výrobců a udržitelné financování nezbytných veřejných služeb.

1.3 Pilíře udržitelného rozvoje

Z definice UR jsou odvozovány tři pilíře udržitelnosti – ekonomický, environmentální a ekonomický. Udržitelný rozvoj spočívá ve vyváženosti všech tří pilířů [16, 18, 25].

Ekonomický pilíř

Ekonomický pilíř je soustředěn především na posilování konkurenceschopnosti ekonomiky za předpokladu respektování limitu environmentálních zdroj. Podstata ekonomicky udržitelného rozvoje spočívá v tom, že není spotřebováváno více, než bylo vyprodukováno, a tím je tedy zachováván úhrn bohatství jako základní kapitál.

Základní kapitál je možné si představit jako souhrn přírodního kapitálu, který vyjadřuje úhrn přírodních zdrojů, vyrobeného kapitálu, tj. výrobků vzniklých lidskou prací, a nepostižitelného kapitálu představovaného lidmi, jejich vzděláním a institucemi. Pro zachování bohatství a tedy také pro udržitelný ekonomický rozvoj je důležité, zda a do jaké míry lze nahradit přírodní kapitál kapitálem vyrobeným nebo nepostižitelným.

Environmentální pilíř

Environmentální pilíř je zaměřen na oblast ochrany životního prostředí, respektive jeho zlepšování a zabránění vyčerpání omezených přírodních zdrojů. Spolu s opatřeními implementovanými v rámci ekonomického pilíře by mělo docházet k motivování ekonomických subjektů zavádět ekologicky šetrné technologie. Kvalitní životní prostředí má potom silné dopady na kvalitu života populace, což již zasahuje do oblasti sociální.

² Na základě soustavného zvyšování vzdělanosti obyvatel všech věkových kategorií.

Sociální pilíř

Sociální pilíř oproti předešlým pilířům je více zaměřován na přítomnost a jeho vazba na generace budoucí je znatelná pouze v případě problematiky stárnutí populace a z toho plynoucího přesunu břemene důchodového systému na tuto generaci. V rámci opatření sociálního pilíře by mělo být sledováno především zajištění kvalitního života z hlediska absence hmotné nouze. V některých případech je do problematiky zajištění kvalitního života přidávána i otázka uspokojování kulturních potřeb.

Sociální rozměr se týká jednak lidí jako jednotlivců a zároveň jako společnosti. Lidský rozvoj znamená odstranění chudoby, zlepšování zdraví, delší průměrný věk, méně nemocí, ale také vzdělanost, slušné životní podmínky, bezpečnost. Společenský rozvoj se týká především institucí demokracie, zabezpečení lidských práva a svobod a spravedlivého společenského uspořádání.

1.4 Principy udržitelného rozvoje

Principy udržitelného rozvoje jsou založeny na přesvědčení, že potřeby současné generace je nutné naplňovat takovým způsobem, aby nebyly oslabovány možnosti příštích generací. Jejich součástí je podpora života v jeho různorodosti, založená na principu demokracie, rovnosti pohlaví, solidarity, respektování právního systému a lidských práv, včetně míru a stejných příležitostí pro všechny.

Udržitelný rozvoj neznamena pouze ochranu životního prostředí a efektivní využívání přírodních zdrojů, ale také zahrnuje aspekty ekonomické a sociální soudržnosti. Principy udržitelného rozvoje by se měly také promítat i do rozhodovacích procesů na všech úrovních. Jednotlivé úrovně veřejné správy a vlády státu by měly spolupracovat, vzájemně se podporovat a navzájem respektovat svá specifika a odlišnosti [4].

Mezi základní principy udržitelného rozvoje je zahrnováno [22]:

- **propojení základních oblastí života** – ekonomické, sociální a životní prostředí (řešení zohledňující pouze jednu nebo dvě z nich není dlouhodobě efektivní);
- **dlouhodobá perspektiva** – každé rozhodnutí je třeba zvažovat z hlediska dlouhodobých dopadů, je třeba strategicky plánovat;

- **kapacita životního prostředí je omezená** – nejenom jako zdroje surovin, látek a funkcí potřebných k životu, ale také jako prostoru pro odpady a znečištění všeho druhu;
- **předběžná opatrnost** – důsledky některých činností nejsou vždy známé, neboť poznání zákonitostí fungujících v životním prostředí je stále ještě na nízkém stupni, a proto je na místě předběžná opatrnost;
- **prevence** - je mnohem efektivnější než následné řešení dopadů, na řešení problémů, které již vzniknou, musí být vynakládáno mnohem větší množství zdrojů, ať už časových, finančních či lidských;
- **kvalita života** - má rozměr nejen materiální, ale také společenský, etický, estetický, duchovní, kulturní a další, lidé mají přirozené právo na kvalitní život;
- **sociální spravedlnost** - příležitostí i zodpovědnosti by měly být děleny mezi země, regiony i mezi rozdílné sociální skupiny (chudoba je ohrožující faktor udržitelného rozvoje);
- **zohlednění vztahu "lokální - globální"** - činnosti na místní úrovni ovlivňují problémy na globální úrovni, vytvářejí je nebo je mohou pomoci řešit a naopak;
- **vnitrogenační a mezigenační odpovědnost (rovnosti práv)** - zabezpečení národnostní, rasové i jiné rovnosti, respektování práv současných i budoucích generací na zdravé životní prostředí a sociální spravedlnost;
- **demokratické procesy** - zapojením veřejnosti již od počáteční fáze plánování vytváříme nejen objektivnější plány, ale také obecnou podporu pro jejich realizaci.

1.5 Indikátory udržitelného rozvoje

Indikátory udržitelného rozvoje jsou vyjadřovány ukazateli vývoje určitého vybraného jevu získané průběžným sledováním, zaznamenáváním a vyhodnocováním souboru přesně stanovených údajů. Vzhledem k tomu, že je řeč o udržitelném rozvoji, jde o sledování jevů, které s tímto tématem úzce souvisí. Ve správně zvoleném indikátoru jsou odrážena hlediska sociální spravedlnosti, zájmy místní ekonomiky a ochrany životního prostředí.

Indikátory umožňují vidět problematické oblasti a ukázat cestu k jejich nápravě. Uplatnění jednotné sady zvolených indikátorů umožňuje posouzení, jak si daný objekt vede ve srovnání s jinými a usnadňuje poznání jejich silných a slabých stránek [23].

Optimálním způsobem používání indikátorů je jejich sdružování do ucelených souborů či sad, například složkově nebo sektorově pojaté soubory. Indikátory mohou být rozlišovány také podle geografické dimenze³ případně podle jiných kritérií. Užití indikátorů nachází své uplatnění zejména při politickém rozhodování na všech úrovních, včetně mezinárodních a pro informování veřejnosti [2].

Mají-li být indikátory použitelné, musí splňovat řadu kritérií. Mezi základní požadavky jsou zařazeny následující kritéria [24]:

- **Správnost** – indikátory musí být správné. Žádná data nejsou naprosto správná, nesmí však být zatíženy významnějšími chybami.
- **Reprezentativnost** – musí být zřejmé, jakým jevem je daný indikátor reprezentován. Kromě volby správného obsahu je předpokládána také volba správné prostorové a časové dimenze ukazatele.
- **Jedinečnost** – získané údaje mají být jedinečné, nemají mít redundantní, opakované, dublovat nějaké již existující informace. Každý indikátor má mít svou specifickou a originalitu a nesmí opakovat to, co je již známo odjinud.
- **Spolehlivost** – data musí být prověřována co do své spolehlivosti, potvrzována několika nezávislými měřeními, případně výsledky získanými zásadně různými metodami. Kontrola a zajištění kvality dat je velmi důležitým oborem.
- **Srovnatelnost** – většina měření, vzorkování či statistických šetření je mezinárodně standardizována. Tato standardizace zaručuje srovnatelnost dat v mezinárodním měřítku a v dlouhém časovém období. Předpokladem pro spolehlivost je však správnost a spolehlivost údajů.
- **Významnost** – indikátory musí být významné v dané souvislosti. Význam indikátoru může být buď v širokém kontextu trvale udržitelného rozvoje a celkového rozvoje společnosti, nebo specifický pro danou složku prostředí či jiný daný jev, jako je například stav ovzduší.

³ Rozdělení indikátorů dle geografické dimenze může být např. na celostátní, regionální či lokální.

- **Měřitelnost, možnost získání dat** – získání podkladových údajů musí být technicky možné. Technická stránka měření a odebírání vzorků je jedno z klíčových záležitostí, které je nutno věnovat pozornost při konstrukci monitorovacích systémů a plánovacích měřících programů.
- **Minimalizace negativních účinků na prostředí** – při vzorkování a měření může někdy docházet k poškozování a dokonce až ke zničení pozorovaného jevu. Zejména při získávání dat o živé přírodě se někdy používají až absurdní postupy, což je potřeba minimalizovat.

1.6 Dílčí závěr

Cílem této kapitoly bylo definování základních pojmů vztahujících se k problematice udržitelného rozvoje. V rámci obsahu kapitoly byla uvedena základní definice, strategie vztahující se k České republice a principy udržitelného rozvoje. Dále byla věnována pozornost oblastem udržitelného rozvoje a bližší specifikace jednotlivých pilířů udržitelného rozvoje. V závěru kapitoly byla popsána problematika indikátorů a kritéria jejich použitelnost.

2 Support Vector Machine

V této kapitole bude pozornost zaměřena na metodu Support Vector Machine, která byla vybrána pro návrh klasifikátoru. V rámci přiblížení této metody bude vysvětlena základní problematika.

2.1 Úvod do problematiky Support Vector Machine

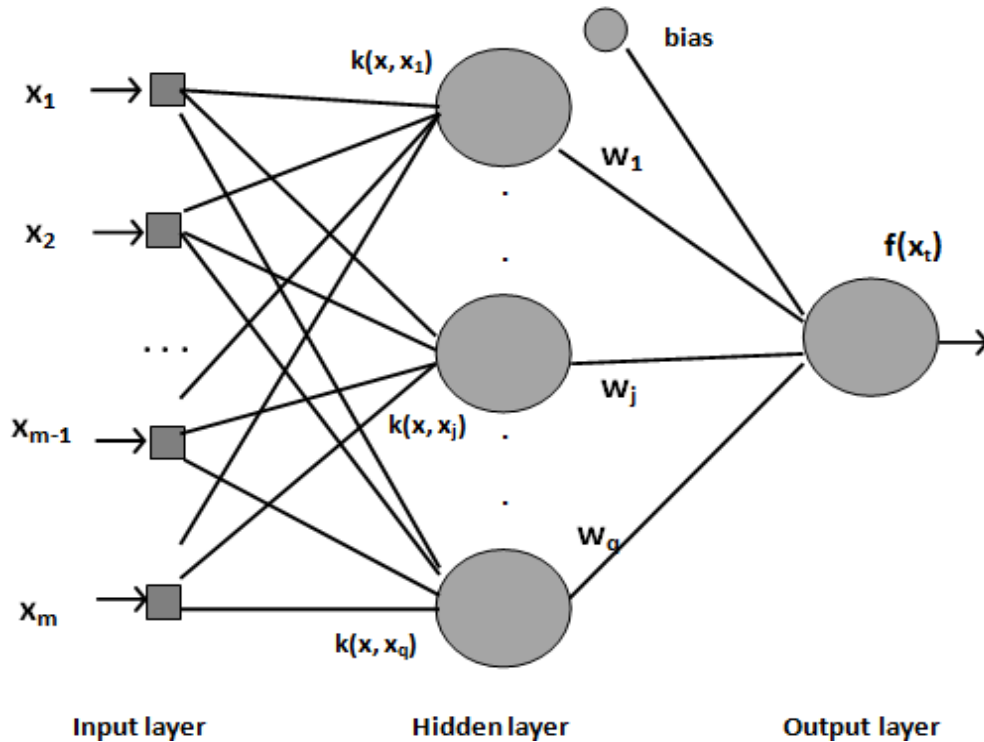
Řada metod strojového učení se vyznačuje jednoduchými a efektivními algoritmy učení, např. jednovrstvé umělé neuronové sítě. Ale z hlediska řešení obecné úlohy najít hranice, které oddělují určité třídy ve vstupním prostoru, jsou velmi silně omezeny schopností naučit se pouze lineární oddělovače (přímky, roviny, nadroviny).

Na druhé straně existují metody jako mnohvrstvé umělé neuronové sítě, které jsou schopny reprezentovat obecné nelineární funkce. Jejich nevýhodou však je často velmi obtížné učení, protože prakticky vždy existuje riziko uvíznutí v lokálním minimu chybové funkce a navíc je učení silně komplikováno hledáním vysokého počtu vah v mnohadimensionálním prostoru.

K alternativním, relativně novým metodám patří Support Vector Machine (podpůrné algoritmy, SVM), které tvoří určitou kategorii tzv. jádrových algoritmů (*kernel machines*). Tyto metody se snaží využít výhody poskytované efektivními algoritmy pro nalezení lineární hranice a zároveň jsou schopny reprezentovat vysoce složité nelineární funkce. Jedním ze základních principů je převod daného původního vstupního prostoru do jiného, vícedimensionálního, kde již lze od sebe oddělit třídy lineárně [10].

SVM je typický algoritmus strojového učení, hledající nadrovinu, která v prostoru příznaků optimálně rozděluje trénovací data. Cílem SVM je nalezení pouze jednoho optimálního lineárního oddělovače. Optimální lineární oddělovač poskytuje co nejširší pásmo mezi ním a pozitivními příklady na jedné straně a negativními na druhé [1].

Obrázek 2.1 znázorňuje architekturu Support Vector Machine [7].



Obrázek 2.1: Architektura Support Vector Machine. Zdroj: [7]

2.2 Optimální nadrovina pro lineárně separovaná data

Nechť je dán trénovací vzorek $\{(x_i, d_i)\}_{i=1}^N$, kde x_i je vstupní vektor pro i -tý příklad a d_i je odpovídající požadovaná odezva neboli cílový výstup. Dále necht' je dáno, že třídy reprezentované podmnožinami $d_i = +1$ a $d_i = -1$ jsou lineárně separované. Rovnice rozhodovací hladiny ve formě nadroviny, která vykonává separaci je dána vztahem

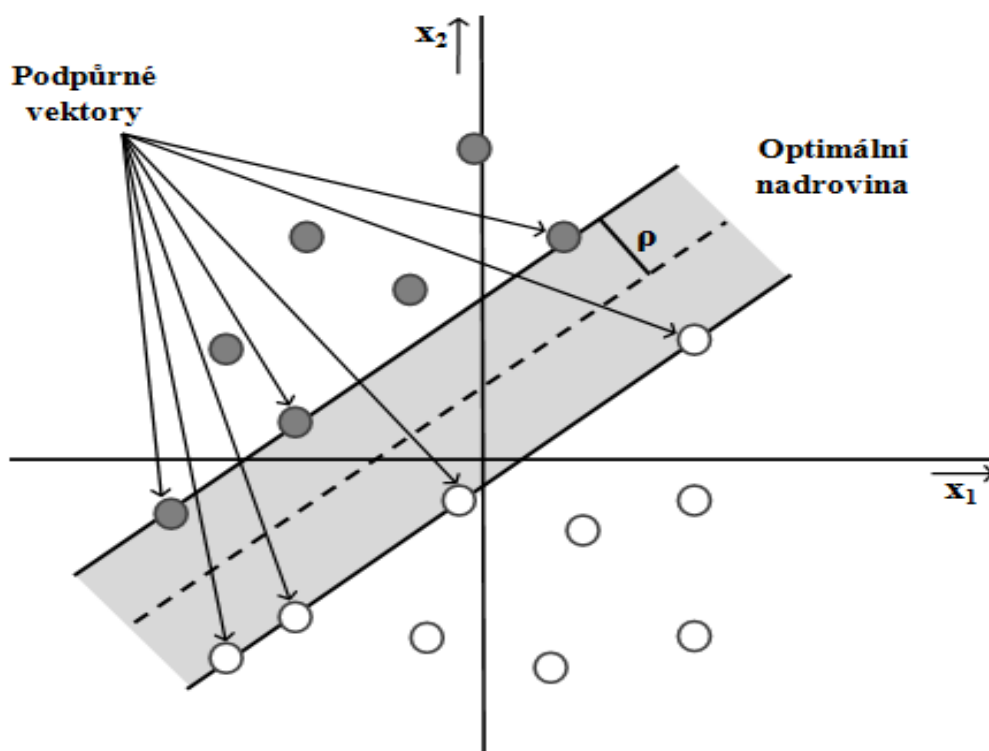
$$w^T x_i + b = 0, \quad (2.1)$$

kde x je vstupní vektor, w je nastavitelný váhový vektor a b vyjadřuje bias neboli odklon či zkreslení. Může být tedy definováno

$$\begin{aligned} w^T x_i + b &\geq 0, \quad \text{pro } d_i = +1, \\ w^T x_i + b &< 0, \quad \text{pro } d_i = -1. \end{aligned} \quad (2.2)$$

Pro daný vektor w a bias b je rozdělení mezi nadrovinou definovanou vztahem (2.1) a nejbližšími datovými body nazýváno Margin rozdělení, označovaný ρ . Cílem SVM je nalézt takovou nadrovinu, pro kterou je Margin rozdělení maximalizováno [8].

Na základě této podmínky je rozhodovací hladina uváděna jako optimální nadrovina. Obrázek 2.2 prezentuje geometrickou konstrukci optimální nadroviny pro dvourozměrný vstupní prostor.



Obrázek 2.2: Optimální nadrovina pro lineárně separovaná data. Zdroj: [8]

Nechť w_0 a p_0 vyjadřuje optimální hodnoty váhového vektoru a bias. Odpovídající optimální nadrovina reprezentující rozhodovací hladinu vstupního prostoru je definována

$$w_0^T x + b_0 = 0. \quad (2.3)$$

Diskriminační funkce

$$g(x) = w_0^T x + b_0 \quad (2.4)$$

vyjadřuje algebraické vyjádření vzdálenosti od x k optimální nadrovině [8].

Otázkou je nalezení parametrů w_0 a p_0 pro optimální nadrovinu danou trénovací množinou $\tau = \{(x_i, d_i)\}$. S ohledem na výsledky zobrazené ve vztahu (2.2) musí páry (w_0, p_0) splnit podmínky

$$\begin{aligned} w_0^T x_i + b_0 &\geq 1, \text{ pro } d_i = +1, \\ w_0^T x_i + b_0 &\leq -1, \text{ pro } d_i = -1. \end{aligned} \quad (2.5)$$

Jednotlivé body (x_i, d_i) , pro které je splněn první nebo druhý řádek vztahu (3.5) jsou Support Vectors. Odtud také pochází název Support Vector Machines. Podpůrné vektory jsou ty datové body, které jsou nejbližší rozhodovací hladině, a proto jsou nejobtížněji klasifikovatelné. Mají přímý vliv na optimální umístění rozhodovací hladiny. Pod optimálním umístěním rozhodovací hladiny se rozumí takové umístění, kdy je Margin rozdělení maximální. Takto umístěná rozhodovací hladina je nazývána optimální rozhodovací nadrovinou [8, 27].

2.3 Optimální nadrovina pro lineárně neseparovaná data

V této fázi je vyjádřen druhý možný a přitom složitější případ dat, která nejsou kompletně separovaná. Pro danou množinu trénovacích dat není možné zkonstruovat oddělující nadrovinu bez výskytu určitých klasifikačních chyb. Cílem je nalézt takovou optimální nadrovinu, která minimalizuje pravděpodobnost klasifikační chyby.

Margin rozdělení mezi třídami je nazýváno Soft Margin, pokud datové body (x_i, d_i) porušují podmínku

$$d_i(w^T x_i + b) \geq 1, \quad \text{pro } i = 1, 2, \dots, N. \quad (2.6)$$

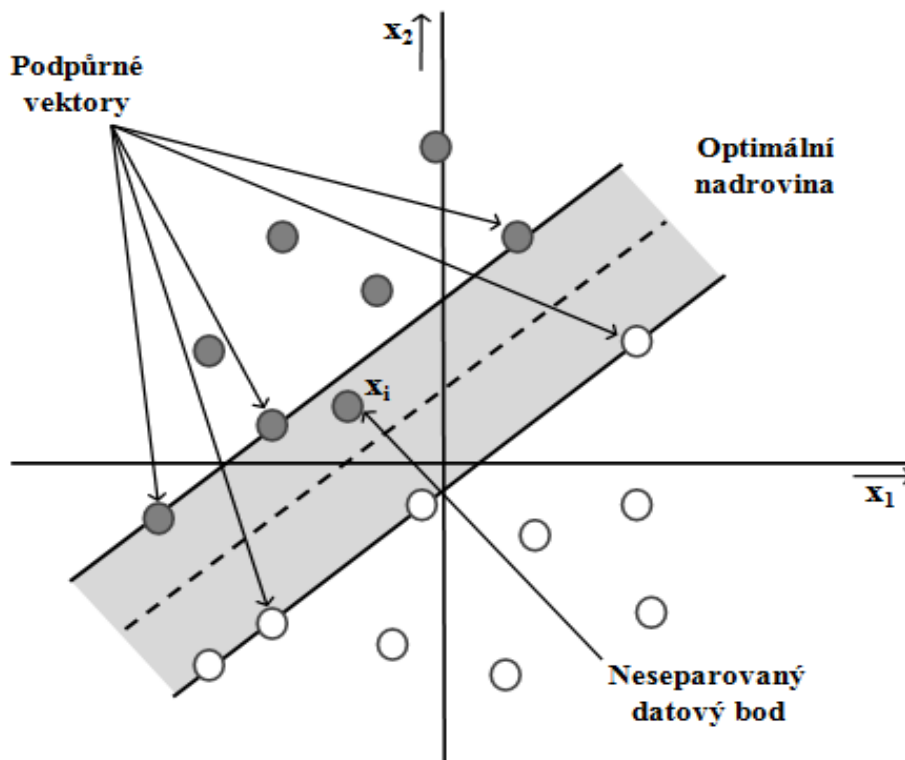
Porušení této podmínky se může vyskytnout v jednom z následujících případů [8]:

- datové body (x_i, d_i) spadají dovnitř oblasti rozdělení, ale na pravou stranu rozhodovací hladiny (viz obrázek 2.3),
- datové body (x_i, d_i) spadají špatnou stranu rozhodovací hladiny (viz obrázek 2.4).

Pro formální ošetření neseparovaných datových bodů je představena nová množina nezáporných skalárních proměnných $\{\xi_i\}_{i=1}^N$. Po úpravě definice rozdělovací nadroviny je zápis následující

$$d_i(w^T x_i + b) \geq 1 - \xi_i, \quad \text{pro } i = 1, 2, \dots, N, \quad (2.7)$$

kde ξ_i se nazývají doplňkové proměnné (Slack Variables). Tyto proměnné měří odchylku datových bodů od ideálních podmínek rozdělení. Pro $0 \leq \xi_i \leq 1$ datové body spadají dovnitř oblasti rozdělení, ale na pravou stranu rozhodovací hladiny (viz obrázek 2.3) [8].



Obrázek 2.3: Optimální nadrovina pro neseparovaná data I. Zdroj: [8]

Pro $\xi_i > 1$ datové body spadají na špatnou stranu rozdělovací nadroviny (viz obrázek 2.4). Podpůrné vektory jsou body, které splňují vztah (2.7), právě tak i v případě, že $\xi_i > 0$. Jsou tedy definovány stejně pro případ jak lineárně separovatelných, tak neseparovatelných dat.

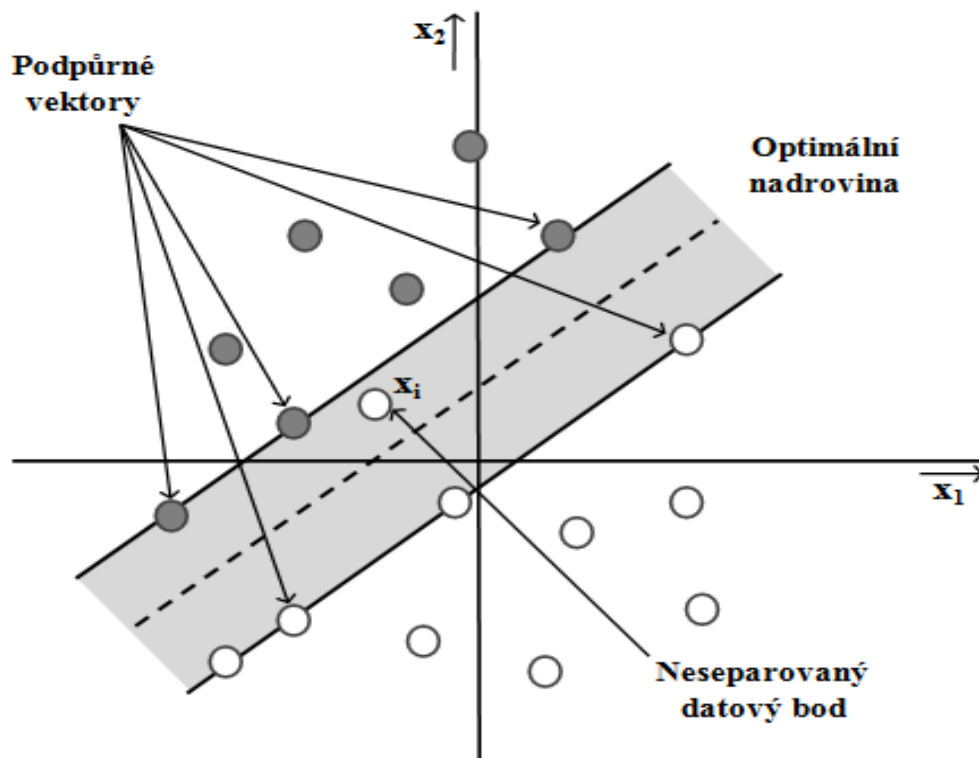
Cílem je nalézt takovou rozdělovací nadrovinu, pro kterou je chyba nesprávné klasifikace minimalizována. To může být vykonáno přes minimalizaci funkce

$$\phi(\xi) = \sum_{i=1}^N I(\xi_i - 1), \quad (2.8)$$

s ohledem na váhový vektor w , vystavený podmínce popsané vztahem (2.7). Funkce $I(\xi)$ je indikační funkce a je definována

$$I(\xi) = \begin{cases} 0 & \text{pro } \xi \leq 0 \\ 1 & \text{pro } \xi > 0 \end{cases}. \quad (2.9)$$

Minimalizace $\phi(\xi)$ s ohledem na w je nekonvexní optimalizační problém [8].



Obrázek 2.4: Optimální nadrovina pro neseparovaná data II. Zdroj: [8]

2.4 Dílčí závěr

Kapitola byla zaměřena na teoretický popis SVM. V první části byla kapitola zaměřena na přiblížení problematiky SVM. Druhá a třetí část kapitoly vyjadřovala formulaci optimální nadroviny pro lineárně separovaná a neseparovaná data. Support Vector Machines jsou používány pro klasifikaci i pro predikci vstupních dat.

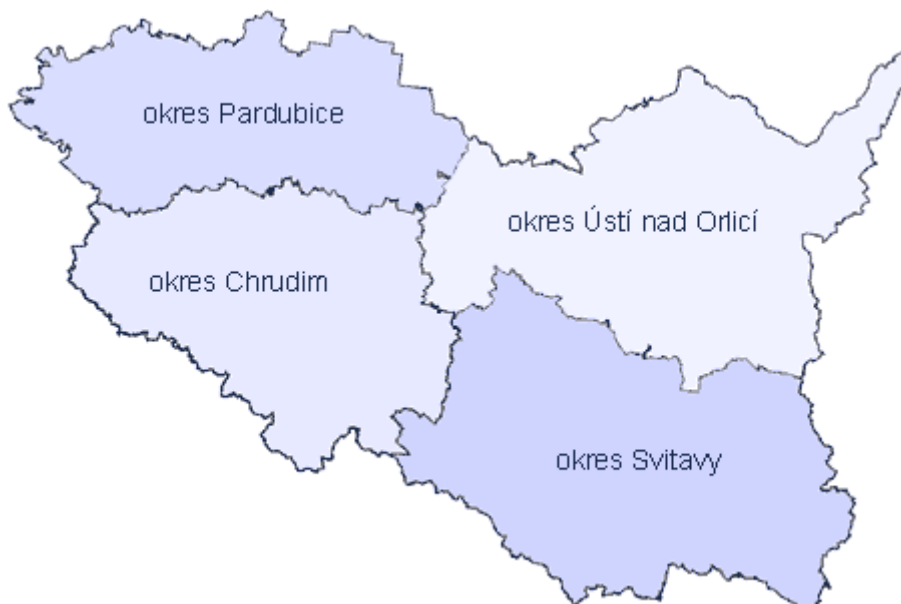
3 Analýza vstupních dat

Kapitola formuluje stručnou charakteristiku Pardubického kraje, k němuž se jednotlivá vstupní data budou vztahovat, a základní indikátory udržitelného rozvoje, které budou výchozí pro následnou klasifikaci. Mezi klíčové indikátory patří především ukazatele z ekonomické, environmentální a sociální oblasti udržitelného rozvoje.

3.1 Pardubický kraj

Pardubický kraj představuje jeden ze čtrnácti krajů České republiky, jehož rozloha přesahuje 4 500 km² a žije v něm přes pět set tisíc obyvatel. Sousedí s Olomouckým, Jihomoravským, Středočeským a Královéhradeckým krajem, dále s krajem Vysočina a na severu s polským Dolnoslezským vojvodstvím.

V kraji se nachází celkem 451 obcí, z nichž 15 jsou obce s rozšířenou působností a 26 obce s pověřeným obecním úřadem. Mezi těmito obcemi se dále vyskytuje 34 měst a 6 městysů. Sídlním městem kraje je statutární město Pardubice. Všechny tyto obce jsou soustředěny ve čtyřech okresech, které jsou znázorněny na obrázku 3.1, a to Pardubice, Chrudim, Ústí nad Orlicí a Svitavy [19].



Obrázek 3.1: Pardubický kraj. Zdroj: [19]

3.2 Indikátory udržitelného rozvoje

Za indikátory udržitelného rozvoje bylo na základě rozhodnutí experta zvoleno celkem 42 ukazatelů z oblasti ekonomické, environmentální a sociální, které umí podat přehled o udržitelném rozvoji jednotlivých obcí. Vybrané ukazatele se vztahují k základním pilířům udržitelného rozvoje a jsou dostupné a měřitelné na úrovni obcí ve zvoleném Pardubickém kraji.

3.2.1 Ekonomické indikátory

Na základě rozhodnutí experta bylo vybráno celkem 11 ekonomických indikátorů, které jsou znázorněny v tabulce 3.1. Každý z indikátorů bude dále blíže specifikován [5, 6, 9].

Podíl ekonomicky aktivních v priméru a sekundéru (p_1)

Parametr vyjadřuje podíl počtu ekonomicky aktivních osob v primárním a sekundárním sektoru hospodářství na celkovém počtu ekonomicky aktivních osob v daném území.

Koncentrace ekonomiky (p_2)

Parametr koncentrace ekonomiky udává koncentraci zaměstnanosti obyvatel v daném odvětví ekonomiky ve vztahu k obci. Koncentraci ekonomiky lze poté vyjádřit jako podíl počtu obyvatel zaměstnaných v daném odvětví ekonomiky vzhledem k celkovému počtu zaměstnaných obyvatel.

Průměrná mzda / průměr kraje (p_3)

Tento parametr vyjadřuje podíl, který je složen z průměrné mzdy v obci a průměrného platu v kraji. Jde o porovnání zvoleného parametru v rámci obce a kraje.

Dluhová služba (p_4)

Ukazatel dluhové služby je jedním z indikátorů, jež sleduje hospodaření obce a její schopnost vypořádat se se svými finančními závazky. Dluhová služba je charakterizována jako roční platby úroků spolu s ročními splátkami. Opakující se příjmy představují celkové příjmy bez jednorázových a kapitálových příjmů. Ukazatel se pohybuje v intervalu od nuly do jedné.

Dluh na obyvatele (p_5)

Dluh na obyvatele představuje zadluženost obce, která je přepočtena na jednoho obyvatele. Hodnotu parametru je nutné porovnávat s ostatními obcemi v daném kraji.

Podíl opakujících se příjmů a běžných výdajů (p₆)

Podíl opakujících se příjmů a běžných výdajů je hodnotícím parametrem, jenž ukazuje na kvalitu rozpočtového hospodaření. Pokud je stabilně $p_6 > 1$, tj. že běžný rozpočet je přebytkový, a zároveň trend ukazuje na rostoucí hodnotu parametru, finanční situace obce je dobrá.

Podíl vlastních příjmů (p₇)

Podíl vlastních příjmů vyjadřuje podíl vlastních příjmů na celkových příjmech. Čím vyšší je tento podíl, tím větší je rozpočtová a daňová nezávislost obce a nižší zadluženost obce.

Podíl kapitálových výdajů (p₈)

Parametr kapitálových výdajů lze vyjádřit jako podíl kapitálových výdajů na celkových výdajích obce. Ukazatel je vyjadřován v intervalu od nuly do jedné. Větší hodnota parametru naznačuje snahu investování obce.

Likvidní majetek na obyvatele (p₉)

Likvidním majetkem jsou často označovány rozlehlé pozemky, komerční budovy, zemědělské pozemky a majetek sloužící podnikatelským účelům v majetku obci. Parametr je vyjádřen jako podíl likvidního majetku vzhledem k počtu obyvatel obce.

Přijaté dotace na obyvatele (p₁₀)

Dotace jsou definovány jako peněžní prostředky státního rozpočtu, státních finančních aktiv či dalších zdrojů poskytnuté obci na stanovený účel. Indikátor vypovídá o aktivitě obci v oblasti získávání finančních prostředků z různých dotačních titulů, které jsou důležité pro financování či spolufinancování různých aktivit v obci. Výsledný parametr je poté vyjádřen jako podíl celkové výše získaných finančních prostředků vzhledem k počtu obyvatel obce.

Počet podnikatelských subjektů / tis. obyvatel (p₁₁)

Počet aktivních podnikatelských subjektů vypovídá o kvalitě podnikatelského prostředí a o ekonomické životaschopnosti obce. Pro intenzitu rozvoje obce je důležité, kolik drobných, středních a velkých podnikatelských subjektů v obci působí. Indikátor zprostředkovaně vypovídá také o množství nabízených služeb či možnostech zaměstnání v obci. Parametr vyjadřuje počet podnikatelských subjektů působících v obci přepočítaný na tisíc obyvatel.

Tabulka 3.1: Souhrn ekonomický indikátorů

Ekonomické indikátory	
p ₁	Podíl ekonomicky aktivních (v priméru a sekundéru, %)
p ₂	Koncentrace ekonomiky
p ₃	Průměrná mzda / průměr kraje
p ₄	Dluhová služba
p ₅	Dluh na obyvatele
p ₆	Podíl opakujících se příjmů a běžných výdajů
p ₇	Podíl vlastních příjmů
p ₈	Podíl kapitálových výdajů
p ₉	Likvidní majetek na obyvatele
p ₁₀	Přijaté dotace na obyvatele
p ₁₁	Počet podnikatelských subjektů / tis. Obyvatel

3.2.2 Environmentální indikátory

Mezi environmentální indikátory bylo zařazeno celkem 11 indikátorů, které se vztahují k životnímu prostředí. Seznam stanovených indikátorů je zobrazen v tabulce 3.2. Jednotlivé indikátory jsou dále blíže specifikovány [5].

Kanalizace s ČOV (p₁₂)

Kanalizace s ČOV představuje ukazatel vyjadřující informaci o existenci kanalizace včetně čističky odpadních vod v obci, který může nabývat pouze dvou stavů. V prvním případě se v dané obci vyskytuje kanalizace s čističkou odpadních vod a tento stav je označen jedničkou. V opačném případě je tato skutečnost označena nulou.

Kanalizace bez ČOV (p₁₃)

Kanalizace bez ČOV je obdobou předchozího indikátoru. Představuje ukazatel vyjadřující informaci o existenci kanalizace bez čističky odpadních vod v obci, který může nabývat také pouze dvou stavů. V případě, že v obci je zavedena kanalizace bez čističky odpadních vod, pak je tento stav označen jedničkou, v opačném případě nulou.

Podíl orné půdy ze zemědělské půdy (p₁₄)

Tento indikátor ukazuje, jak velkou část zemědělské půdy v procentech vyjadřuje půda orná na daném území.

Podíl zastavěných a ostatních ploch z celkové výměry (p₁₅)

Indikátor popisuje, jak velkou část celkové výměry území v procentech vyjadřuje součet výměr zastavěných a ostatních ploch.

Podíl vodních ploch z celkové výměry (p₁₆)

Indikátor popisuje, jak velkou část celkové výměry území v procentech vyjadřují vodní plochy.

Koeficient ekologické stability (p₁₇)

Koeficient ekologické stability představuje podíl výměr druhů pozemků v daném území. Tento podíl je sestaven jako součet výměr chmelnic, vinic, zahrad, ovocných sadů, trvalých travních porostů, lesní půdy a vodních ploch vzhledem k součtu výměr orné půdy, zastavěné půdy a ostatních ploch.

Zahrady na obyvatele (p₁₈)

Zahrady jsou popisovány jako pozemky zpravidla oplocené, na kterých se převážně a trvale pěstuje zelenina, květiny a jiné zahradní plodiny, a to zpravidla pro vlastní potřebu. Dále také souvislé pozemky osázené ovocnými stromy nebo keři až do výměry 0,25 ha, školky ovocných nebo okrasných stromů, viničné školky, pařeniště a skleníky.

Ovocné sady na obyvatele (p₁₉)

Ovocné sady jsou popisovány jako pozemky o výměře větší jak 0,25 ha osázené ovocnými stromy v hustotě na 1 ha v počtu dle daných rozměrů stromů. Patří sem též pozemky, kde se pěstují výhradně zákrsky ovocných stromů v hustotě nejméně 500 zákrsků na 1 ha a pozemky kde se pěstuje výhradně černý rybíz v hustotě 1000 keřů na 1 ha nebo ostatní druhy rybízu nebo angreštu a hustotě nejméně 2000 keřů na 1 ha. Ukazatel poté vyjadřuje počet hektarů na jednoho obyvatele.

Trvalé travní porosty (p₂₀)

Do ukazatele trvalé travní porosty je zařazován dřívější ukazatel Louky, které jsou popisovány jako pozemky porostlé travinami, u nichž hlavním výtěžkem je tráva, i když se nahodile spásají. Dále sem patří Pastviny, které jsou popisovány jako pozemky porostlé travinami, které jsou určeny k trvalému spásání, a také pastevní výběhy pro skot, vepře a drůbež. Ukazatel vyjadřuje počet hektarů na jednoho obyvatele.

Lesní půda na obyvatele (p₂₁)

Ukazatel vyjadřuje počet hektarů lesní půdy na jednoho obyvatele. Součástí lesní půdy je půda porostní (půda využívaná přímo k lesní produkci), půda skutečně zalesněná nebo dočasně odlesněná s úmyslem opětovné obnovy lesního porostu, bezlesí (dočasně odlesněná část lesní půdy) a odňaté pozemky zemědělskému půdnímu fondu přidělené lesnímu hospodářství k zalesnění.

Zemědělská půda na obyvatele (p₂₂)

Zemědělská půda vyjadřuje součtový ukazatel udávající souhrn výměr druhů pozemků sloužících bezprostředně zemědělskému výrobnímu procesu jako základní prostředek, z něhož se získává rostlinná produkce. Ukazatel je tvořen jako podíl součtu výměr orné půdy, chmelnic, vinic, zahrad, ovocných sadů trvale travních porostů vzhledem k celkovému počtu obyvatel daného území.

Tabulka 3.2: Souhrn environmentálních indikátorů

Environmentální indikátory	
p ₁₂	Kanalizace s ČOV
p ₁₃	Kanalizace bez ČOV
p ₁₄	Podíl orné půdy ze zemědělské půdy (%)
p ₁₅	Podíl zastavěných a ostatních ploch z celkové výměry (%)
p ₁₆	Podíl vodních ploch z celkové výměry (%)
p ₁₇	Koeficient ekologické stability
p ₁₈	Zahrady na obyvatele (ha/obyv.)
p ₁₉	Ovocné sady na obyvatele (ha/obyv.)
p ₂₀	Trvalé travní porosty (ha/obyv.)
p ₂₁	Lesní půda na obyvatele (ha/obyv.)
p ₂₂	Zemědělská půda na obyvatele (ha/obyv.)

3.2.3 Sociální indikátory

Poslední skupinou navržených indikátorů jsou indikátory sociální. Do této skupiny bylo zařazeno celkem 20 parametrů vztahujících se k sociálním podmínkám a možnostem obyvatel obcí [5, 15].

Vyjíždějící do zaměstnání na počet obyvatel (p₂₃)

Vyjíždějící do zaměstnávání představují osoby, které mají v obci trvalý pobyt, avšak jejich pracoviště je v jiné obci. Těmito osobami mohou být zaměstnanci, zaměstnavatelé, samostatně činní, pracující důchodci a ženy na mateřské dovolené v trvání 28 nebo 37 týdnů. Do této skupiny nelze zahrnout pracující studenty a učně. Následné vyjádření identifikátoru je vztaženo k celkovému počtu obyvatel v dané obci. Výsledku je možno dosáhnout jako podílu osob vyjíždějících do zaměstnání k celkovému počtu obyvatel obce.

Vyjíždějící do škol na počet obyvatel (p₂₄)

Mezi vyjíždějící osoby do škol mimo obec jsou zahrnováni žáci, studenti a učni, kteří mají v obci trvalý pobyt, ale jejich škola je v jiné obci či. Přestože existuje mnoho případů, pracujících studentů a učňů, jsou vždy zařazeni do skupiny vyjížděky do škol, jelikož stále probíhá příprava na budoucí povolání. Podobně jako u vyjíždějících do zaměstnání je potřebných údajů dosaženo pomocí podílu osob vyjíždějících do škol k celkovému počtu obyvatel obce.

Dojíždějící do zaměstnání na počet obyvatel (p₂₅)

Dojíždějící do zaměstnávání jsou vyznačovány osobami, které nemají v obci trvalý pobyt, avšak dojíždějí do obce za prací. Těmito osobami mohou být zaměstnanci, zaměstnavatelé, samostatně činní, pracující důchodci a ženy na mateřské dovolené v trvání 28 nebo 37 týdnů. Indikátor vyjadřuje podíl dojíždějících do zaměstnání vzhledem k celkovému počtu obyvatel obce. Do této skupiny nelze zahrnovat pracující studenty a učně.

Dojíždějící do škol na počet obyvatel (p₂₆)

Dojíždějící osoby do škol jsou představováni žáky, studenty a učni, kteří v dané obci navštěvují školu, avšak jejich trvalé bydliště je v jiné obci. Podobně jako v případě vyjíždějících do škol jsou i do dojíždějících do škol zahrnuti jak pracující studenty a učni, tak i osobami ekonomicky neaktivními, které představují ostatní děti, studenti a učni. Ke stanovení indikátoru pak dochází vypočítáním podílu dojíždějících do škol k celkovému počtu obyvatel obce.

Počet trvale obydlených bytů na počet obyvatel (p₂₇)

Trvale obydlený byt je možné vyjádřit, jako byt, v němž je k trvalému pobytu hlášena alespoň jedna osoba. Dle rozhodnutí stavebního úřadu je možné byt popsat jako soubor místností, popřípadě jednotlivou obytnou místnost, jenž svým stavebně technickým uspořádáním a vybavením splňuje požadavky na trvalé bydlení. Dále zahrnuje vlastní uzavíratelný vstup, prostor pro vaření, pro tělesnou hygienu a obytný prostor. Tento indikátor poté představuje podíl celkového počtu trvale obydlených bytů k celkovému počtu obyvatel obce.

Počet bytů k rekreaci na počet obyvatel (p₂₈)

Byty určené k rekreaci jsou označovány jako byty, v nichž není přihlášena žádná osoba k trvalému ani přechodnému pobytu. Všechny případy, které nespadají do trvale obydlených bytů, což znamená, že v nich není přihlášena žádná osoba k trvalému pobytu, spadají do skupiny neobydlených bytů. Výsledný indikátor je možné vyjádřit jako podíl počtu bytů k rekreaci v obci k celkovému počtu obyvatel.

Podíl obyvatel s pitnou vodou (p₂₉)

Podíl obyvatel s pitnou vodou vyjadřuje podíl počtu obyvatel podle druhu bydlení a vybavení bytu na celkovém počtu obyvatel s tímto druhem bydlení v území. V tomto případě je druh bydlení přestavován trvale obydlenými byty a za vybavení je považováno zásobování pitnou vodou z vodovodu. Celkový indikátor poté vyjadřuje podíl počtu obyvatel žijících v trvale obydlených bytech zásobovaných pitnou vodou z vodovodu vzhledem k celkovému počtu obyvatel obce, žijících v trvale obydlených bytech.

Podíl obyvatel s plynem (p₃₀)

Stejně jako u předchozího indikátoru jde o vyjádření podílu počtu obyvatel podle druhu bydlení a vybavení bytu na celkovém počtu obyvatel s tímto druhem bydlení v území. Celkový indikátor vyjadřuje podíl počtu obyvatel žijících v trvale obydlených bytech, do jejichž bytu byl zaveden plyn, vzhledem k celkovému počtu obyvatel obce, žijících v trvale obydlených bytech.

Podíl obyvatel napojených na kanalizaci (p₃₁)

Podíl obyvatel napojených na kanalizaci představuje další s indikátorů spojený druhem a vybavením bytu. V tomto případě se jedná o trvale obydlené byty, které jsou napojeny na kanalizaci. Podíl poté vyjadřuje počet obyvatel žijících trvale v bytech napojených na kanalizaci vzhledem k celkovému počtu obyvatel žijících v trvale obydlených bytech.

Počet obyvatel (p₃₂)

Počet obyvatel představuje jeden ze základních indikátorů, který udává počet obyvatel území k určitému okamžiku. Do počtu obyvatel jednotlivých obcí jsou zahrnovány osoby s trvalým či dlouhodobým pobytem v obci a to bez ohledu na státní občanství. Zahrnováni jsou také cizinci s trvalým i přechodným pobytem a cizinci, kterým byl přiznán azyl.

Míra nezaměstnanosti (p₃₃)

Míra nezaměstnanosti vyjadřuje jeden ze základních ukazatelů vývoje ekonomiky a je sestavována jako podíl počtu nezaměstnaných registrovaných úřady práce na počtu ekonomicky aktivních osob. Je-li nízká míra nezaměstnanosti, pak je možné usuzovat dobré ekonomické zdraví. V opačném případě pokud je vysoká míra nezaměstnanosti, dochází ke zvyšování výdajů na sociální služby.

Podíl VŠ obyvatel (p₃₄)

Podíl vysokoškolsky vzdělaných obyvatel je vyjadřován jako poměr počtu obyvatel, kteří dosáhli vysokoškolského vzdělání, a celkového počtu obyvatel dané obce. Vzdělanost obyvatel má do značné míry veliký vliv na ekonomický rozvoj obcí.

Stanice vlaku (p₃₅)

Stanice vlaku je ukazatelem vyjadřující informaci o umístění vlakové stanice v obci. V datech může docházet pouze ke dvěma možným stavům. Vyjádření stavu nulou znamená, že v dané obci se nevyskytuje žádná vlaková stanice. V opačném případě, kdy obec má alespoň jednu vlakovou stanici, je stav vyjádřen jedničkou.

Přirozený přírůstek (p₃₆)

Indikátor přirozený přírůstek je vyjádřen jako rozdíl počtu živě narozených dětí a počtu zemřelých obyvatel na daném území za stejné období, kterým obvykle bývá jeden rok. Velkou roli u tohoto indikátoru hraje porodnost, která je ovlivněna mnoha faktory jako například ekonomickou situací či upřednostňováním kariéry před rodinným zázemím.

Saldo migrace (p₃₇)

Saldo migrace vyjadřuje podobně jako přirozený přírůstek srovnání dvou protipólů v určitém časovém horizontu. V tomto případě se jedná o rozdíl počtu přistěhovalých a vystěhovalých osob v daném území za stejné časové období, kterým je obvykle jeden rok. Tento indikátor je ovlivněn především trhem práce, který nabízí lidem nové možnosti. Saldo může být ovlivněno jak pozitivními tak i negativními vlivy. Jedním z největších vlivů je bezesporu nezaměstnanost. V oblastech s vysokou nezaměstnaností bude pravděpodobně docházet k odchodu lidí do jiných oblastí za prací a naopak.

Počet obyvatel ve věku 0 – 14 na celkovém počtu obyvatel (p₃₈)

Tento indikátor vyjadřuje podíl počtu obyvatel podle věku na celkovém počtu obyvatel dané obce. Přesněji lze říci, že vyjadřuje, jak velkou část obyvatel tvoří osoby, které nedosáhli věku 15 let, vzhledem k celkovému počtu obyvatel na daném území.

Počet obyvatel ve věku 65 a více let na celkovém počtu obyvatel (p₃₉)

Tento indikátor představuje stejně jako v předchozím případě podíl počtu obyvatel podle věku na celkovém počtu obyvatel dané obce. V tomto případě jde o vyjádření velikosti skupiny osob, které dosáhli věku 65 let nebo vyššího, vzhledem k celkovému počtu obyvatel na daném území.

Počet uchazečů o zaměstnání na počet obyvatel (p₄₀)

Indikátor představuje podíl celkového počtu uchazečů o zaměstnání evidovaných na úřadu práce s bydlištěm v příslušné obci vzhledem k celkovému počtu obyvatel dané obce. U státního občana České republiky se bydlištěm rozumí adresa místa trvalého pobytu na území České republiky. U cizince, který je občanem Evropské unie nebo jeho rodinným příslušníkem, se bydlištěm rozumí adresa trvalého nebo přechodného pobytu na území České republiky a pokud takový pobyt nemá, adresa místa, kde se na území České republiky obvykle zdržuje. U cizince, který není občanem Evropské unie nebo jeho rodinným příslušníkem, se bydlištěm rozumí adresa místa trvalého pobytu na území České republiky.

Počet dokončených bytů na počet obyvatel (p₄₁)

Počet dokončených bytů vyjadřuje počet bytů v budovách pro bydlení, nových i stávajících, jejichž výstavba byla ve sledovaném období v daném území dokončena a vydaná kolaudační rozhodnutí nabyta právní moci. Mohou jimi být byty v nové výstavbě, nástavbě, přístavbě, přestavbě, dokončené modernizaci či rekonstrukci.

Počet lázeňských léčeben (p₄₂)

Indikátor vyjadřuje celkový počet lázeňských léčeben v dané oblasti. Lázeňská léčebna poskytuje nemocným speciálně zaměřenou ústavní lázeňskou péči, která navazuje na péči poskytovanou v jiných zařízeních léčebně preventivní péče a při níž se využívá především přírodních léčivých zdrojů nebo klimatických podmínek.

Tabulka 3.3: Souhrn sociálních indikátorů

Sociální indikátory	
p ₂₃	Vyjíždějící do zaměstnání na počet obyvatel
p ₂₄	Vyjíždějící do škol na počet obyvatel
p ₂₅	Dojíždějící do zaměstnání na počet obyvatel
p ₂₆	Dojíždějící do škol na počet obyvatel
P ₂₇	Počet trvale obydlených bytů na počet obyvatel
p ₂₈	Počet bytů k rekreaci na počet obyvatel
p ₂₉	Podíl obyvatel s pitnou vodou
p ₃₀	Podíl obyvatel s plynem
p ₃₁	Podíl obyvatel napojených na kanalizaci
p ₃₂	Počet obyvatel
p ₃₃	Míra nezaměstnanosti
p ₃₄	Podíl VŠ obyvatel
p ₃₅	Stanice vlaku
p ₃₆	Přirozený přírůstek
p ₃₇	Saldo migrace
p ₃₈	Podíl obyvatel ve věku 0-14 na celkovém počtu obyvatel (%)
p ₃₉	Podíl obyvatel ve věku 65 a více let na celkovém počtu obyvatel (%)
p ₄₀	Počet uchazečů o zaměstnání na počet obyvatel
p ₄₁	Počet dokončených bytů na počet obyvatel
p ₄₂	Počet lázeňských léčeben

3.3 Dílčí závěr

Tato kapitola je věnována analýze vstupních dat udržitelného rozvoje na úrovni obcí Pardubického kraje. První část obsahuje stručnou charakteristiku Pardubického kraje, k němuž se jednotlivá data vztahují. V další části je vytvořen přehled ekonomických, environmentálních a sociálních indikátorů, které budou sloužit k návrhu modelu klasifikátoru udržitelného rozvoje, a jejich bližší specifikace.

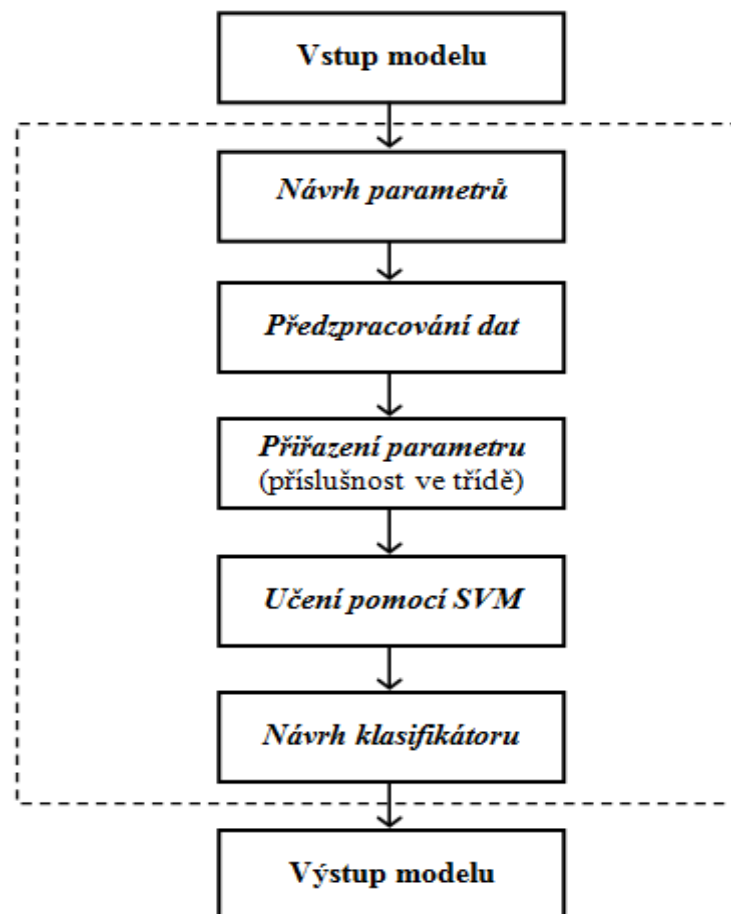
4 Návrh modelu klasifikátoru udržitelného rozvoje

Cílem této kapitoly je navrhnout takový model, který by mohl být využit pro modelování udržitelného rozvoje. Návrh modelu bude sestaven na bázi metody učení s učitelem. Předpoklad této metody tvoří skutečnost, že každému objektu ze vstupních dat bude patřit právě jeden výstupní vektor. Výstupem této kapitoly je navržený model v programovém prostředí Weka, které je volně dostupné.

Další součástí této kapitoly je formulace parametrů metod SVM, charakteristiku vstupních dat a jejich následné rozdělení na testovací a trénovací množinu.

4.1 Návrh modelu pro klasifikaci obcí

Postup při návrhu modelu pro klasifikaci parametrů udržitelného rozvoje obcí je znázorněn na obr. 4.1.



Obrázek 4.1: Návrh modelu pro klasifikaci obcí. Zdroj: [vlastní]

Návrh modelu pro klasifikaci je založen na bázi učení s učitelem. Na základě tohoto učení je znám parametr, který udává příslušnost objektu v dané třídě. Model je složen z návrhu parametrů, předzpracování dat, přiřazení parametru určujícího příslušnost v dané třídě a učení v programovém prostředí.

Vstup do modelu je představován informacemi o obcích z oblasti ekonomické, environmentální a sociální, které slouží ke stanovení hodnot jednotlivých parametrů určených pro následnou klasifikaci. Následně je poté vytvořen vstupní vektor \vec{p} parametrů udržitelného rozvoje obcí.

Následnou částí po stanovení jednotlivých parametrů je fáze předzpracování dat. Tato fáze zahrnuje několik dílčích částí. Mezi nejdůležitější části této fáze bezesporu patří standardizace dat a jejich následné rozdělení na testovací a trénovací množinu. Více k této problematice bude popsáno v kapitole 4.4.

Ve chvíli kdy dochází k ukončení předzpracování dat je nutností přiřadit objektu parametr, jehož účelem je zařazení objektu do jedné ze stanovených tříd. Tento parametr je objektům přiřazen na základě předchozích stanovení experta.

Před závěrečným navržením klasifikátoru, který poskytne nejlepší výsledky, dochází k samotnému učení na bázi učení s učitelem. Toto učení je založeno na algoritmu SVM. Výsledkem učení je navržený klasifikátor, který zařadí obce do zvolených skupin s největší možnou přesností.

4.2 Charakteristika vstupních dat

Vstupní data jsou vyjádřena vstupní datovou maticí, která je tvořena daným počtem obcí v Pardubickém kraji a statistickými ukazateli, které vyjadřují jednotlivé parametry udržitelného rozvoje z oblasti ekonomické, environmentální a sociální. V Pardubickém kraji je zaznamenáno celkem 451 obcí. Na základě rozhodnutí experta bylo vybráno celkem 42 parametrů, z nichž 11 parametrů bylo z oblasti ekonomické, 11 z oblasti environmentální a 20 z oblasti sociální. Tato problematika je podrobněji popsána v kapitole 3.

Vstupní vektor \vec{p} je tvořen pomocí parametrů udržitelného rozvoje obcí a je možné jej zapsat ve tvaru

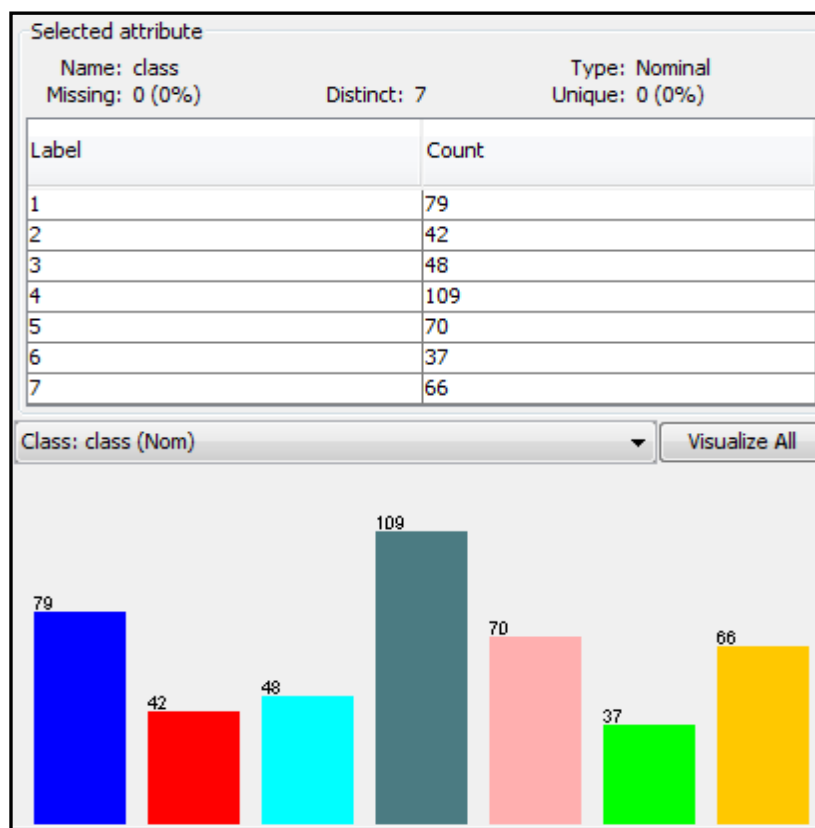
$$\vec{p} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n), n = 42. \quad (4.1)$$

Jednotlivé hodnoty n vyjadřují celkový počet zvolených parametrů.

Vzhledem ke skutečnosti, že metoda SVM je založena na bázi učení s učitelem, je nutností k danému vstupnímu vektoru doplnit parametr t , na základě kterého je možné zjistit, do které ze sedmi tříd byl objekt zařazen. Poté je možno vstupní vektor vyjádřit ve tvaru [27]

$$\vec{p} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n, \mathbf{t}), t = 7. \quad (4.2)$$

Podrobnější přehled o rozdělení obcí do jednotlivých tříd je možné vidět na obrázku 4.2, který zobrazuje počty objektů v jednotlivých třídách.



Obrázek 4.2: Rozdělení objektů do stanovených tříd. Zdroj: [vlastní]

Na základě přiřazených tříd od experta budou data posuzována v oblastech, ke kterým se rámcově váží. Dochází tedy ke vzniku tří vstupních souborů, ve kterých budou jednotlivé oblasti posuzovány zvlášť.

4.3 Předzpracování dat

Před započítím prvních kroků předzpracování dat, je dobré pro lepší orientaci v datech zpracovat popisnou statistiku, která podává stručný přehled o datech v souboru. Z tohoto důvodu byly vytvořeny pro ekonomická, environmentální i sociální data jednotlivé popisné statistiky, které jsou znázorněny v tabulkách v příloze 1.

Hodnoty vstupních parametrů se výrazně liší a jsou vyjádřeny v různých jednotkách. Tato skutečnost může způsobovat, že určité znaky se jeví jako velmi dominantní a jiné naopak málo ovlivňují celkový pohled na data. Z tohoto důvodu je vhodné data upravit tak, aby všechny znaky byly souměřitelné, čehož je možné dosáhnout pomocí standardizace dat. Standardizované hodnoty jsou zpracovány pomocí vztahu

$$\mathbf{x}_{ij} = \frac{z_{ij} - \bar{z}_j}{s_j}, \quad (4.3)$$

kde: - x_{ij} vyjadřuje standardizovanou hodnotu pro i -tý O_i a j -tou proměnou z_j ,

- z_{ij} představuje původní hodnotu pro i -tý objekt O_i a j -tou proměnou z_j ,

- \bar{z}_j je průměr j -té proměnné z_j ,

- s_j vyjadřuje směrodatnou odchylku j -té proměnné z_j .

Po úpravě dat dle stanoveného vztahu mají standardizované hodnoty průměrnou hodnotu rovnu nule a směrodatnou odchylku rovnu jedné [12].

4.4 Rozdělení dat na testovací a trénovací množinu

Jeden z nejdůležitějších kroků předzpracování dat představuje rozklad množiny dat A na množinu testovací a trénovací. Necht' [14] je dán jednoduchý způsob rozkladu množiny objektů A na množinu testovací a trénovací $A = A_{train} \cup A_{test}$ a necht' je známá metoda shlukové analýzy, na základě které dojde k rozložení množiny A na disjunktí podmnožiny – shluky, které obsahují podobné objekty

$$A = C_1 \cup C_2 \cup \dots \cup C_p \quad (4.4)$$

kde i -tý shluk C_i obsahuje n_i objektů z množiny A

$$C_i = \{o_1^{(i)}, o_2^{(i)}, \dots, o_{n_i}^{(i)}\} \subset A \quad (4.5)$$

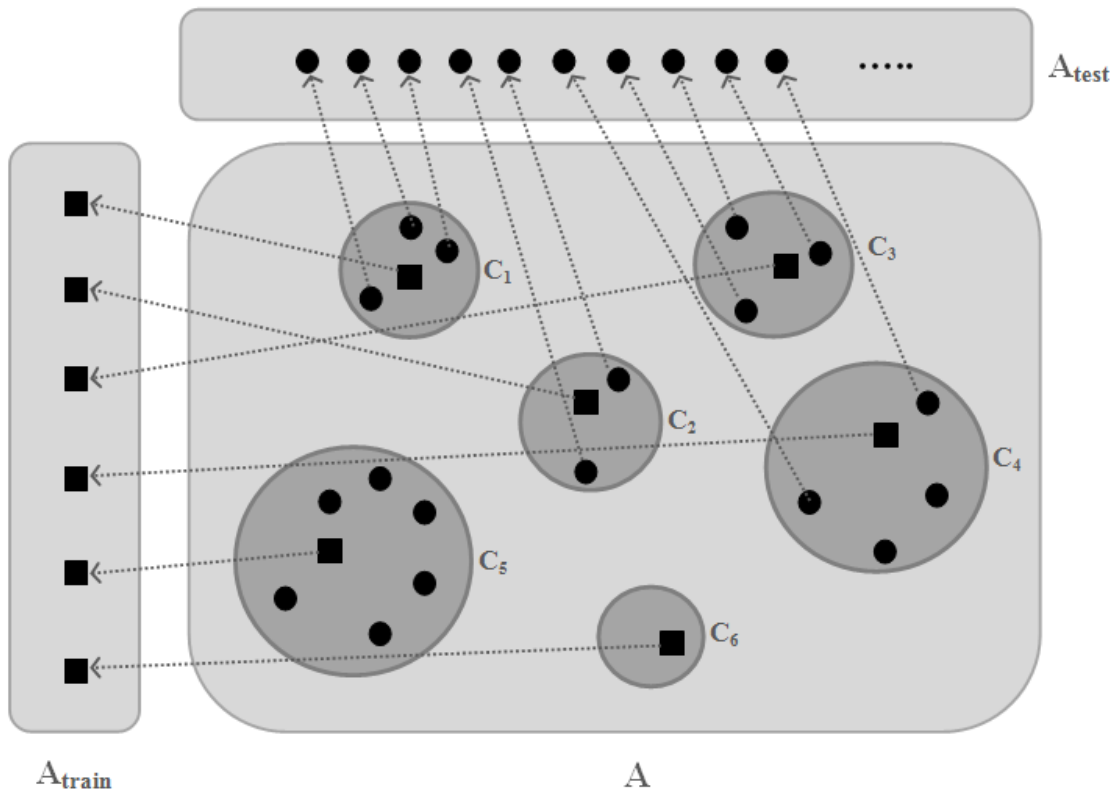
přičemž se předpokládá, že objekt $o_1^{(i)} \in C_i$ je ten objekt z i -tého shluku C_i , který leží nejbližně jeho centru. Tento objekt bude sloužit jako reprezentant objektů ve shluku C_i . Poté je testovací a trénovací množina určena objekty

$$A_{train} = \{o_1^{(1)}, o_1^{(2)}, \dots, o_1^{(p)}\} \quad (4.6)$$

$$A_{test} = (C_1 - \{o_1^{(1)}\}) \cup (C_2 - \{o_1^{(2)}\}) \cup \dots \cup (C_p - \{o_1^{(p)}\}) \quad (4.7)$$

Z vyjádřených vztahů lze říci, že trénovací množina je složena ze všech reprezentantů daných shluků a testovací množina obsahuje zbylé objekty. Počet objektů v trénovací množině je totožný s počtem shluků $|A_{train}| = p$ a $|A_{test}| = |A| - p$.

Schematické znázornění rozkladu objektů množiny A na trénovací a testovací množinu pomocí rozkladu A na shluky C_1, C_2, \dots, C_p je znázorněno na obrázku 4.3. Objekty vyjádřené plnými čtverci představují objekty, které leží nejbližně ke středům příslušných shluků, ostatní objekty jsou vyjádřeny plnými kruhy [14].



Obrázek 4.3: Rozklad objektů na množinu testovací a trénovací. Zdroj: [14]

4.5 Algoritmus Sequential Minimal Optimization

Pro návrh modelu klasifikátoru udržitelného rozvoje bude využito metody SVM. Jako klasifikační algoritmus byl vybrán algoritmus Sequential Minimal Optimization (SMO).

Algoritmus Sequential Minimal Optimization (dále jen SMO) je řazen mezi nové řešení pro trénování SVM. Trénování SVM vyžaduje řešení velkého kvadratického programování (KP) problému optimalizace. Algoritmus SMO se snaží řešení velkého KP rozdělit do série menších KP problémů. Tyto malé KP problémy jsou řešeny analyticky, kdy se snaží o vyhnutí časově náročné číslkové KP optimalizace[1].

SMO je jednoduchý algoritmus, který může rychle řešit SVM KP problém bez jakékoliv zvláštní maticové paměti a bez použití číslkové KP optimalizace. Algoritmus rozebírá celkový KP problém do KP dílčích problémů [20].

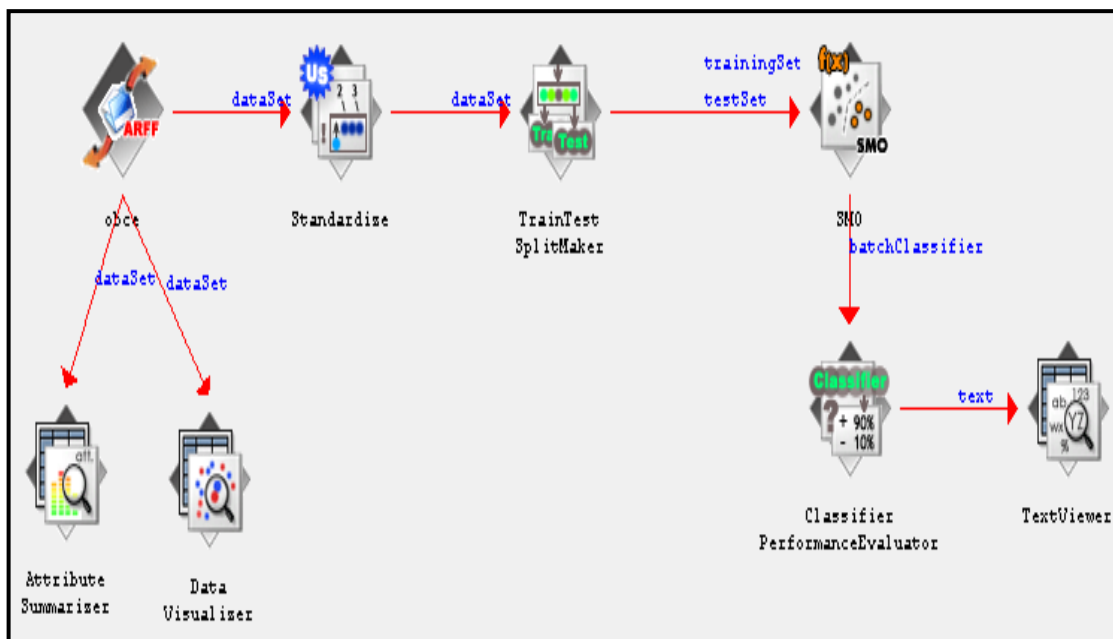
Algoritmus SMO patří mezi učící se funkce (Learner Function). Tyto funkce obsahují několik parametrů, které je nutné na počátku nastavit na optimální hodnoty. V případě nastavení optimálních hodnot dosáhne učící se algoritmus co nejlepších výsledků.

Mezi nejdůležitější parametry algoritmu SMO, které nejvíce ovlivňují kvalitu výsledků, jsou řazeny následující parametry [20]:

- **C – the complexity constant**, která nabývá hodnot $(-\infty ; +\infty)$,
- **N – filter Type**, který popisuje, jakou mají mít data formu (standardizovaná, normalizovaná či beze změny),
- **L – the tolerance parameter**, která nabývá hodnot $(-\infty ; +\infty)$,
- **useRBF – Radial Basic Function**, která nabývá hodnot (True, False).

4.6 Navržený model v programovém prostředí Weka

V programovém prostředí Weka byly navrženy tři modely klasifikace udržitelného rozvoje na základě ekonomických, environmentálních a sociálních indikátorů. Obrázek 4.4 prezentuje návrh modelu ve zvoleném prostředí. Modely byly sestaveny na základě návrhu modelu v podkapitole 4.1. Pro jednotlivé modely bylo zapotřebí nastavení parametrů algoritmu SMO tak, aby výsledky udávali co nejlepší přesnost klasifikace. V rámci výsledků je znázorněno také zařazení objektů do jednotlivých tříd.



Obrázek 4.4: Návrh modelu v prostředí Weka. Zdroj: [vlastní]

4.7 Dílčí závěr

Tato kapitola byla zaměřena na návrh modelu klasifikátoru udržitelného rozvoje a následně jeho sestavení ve zvoleném programovém prostředí Weka. Bylo provedeno předzpracování vstupních dat a jejich následné rozdělení na testovací a trénovací množinu. Pro jednotlivé oblasti parametrů byly nastaveny parametry algoritmu SMO, na základě kterého byly dále zjišťovány přesnosti klasifikace. Po nastavení parametrů na optimální hodnoty byl sestaven ekonomický, environmentální a sociální model klasifikátoru.

5 Analýza navrženého modelu

Analýza navrženého modelu je provedena v programovém prostředí Weka, které je volně dostupné. Veškeré simulace byly provedeny v tomto prostředí. Cílem kapitoly je v průběhu verifikace testovat jednotlivé navržené struktury modelu klasifikace a výběr nejvhodnější z nich. Důležitým krokem je návrh parametrů, které ovlivňují navržené modely a sestavit analýzu klasifikátoru nejvhodnějšího modelu. Další součástí této kapitoly je formulace parametrů metod SVM, charakteristiku vstupních dat a jejich následné rozdělení na testovací a trénovací množinu.

5.1 Programové prostředí Weka

System Weka byl navržen na Novozélandské univerzitě Waikato a představuje jeden z nejrozšířenějších nekomerčních systémů, které jsou vhodné pro dobývání znalostí z dat. V případě potřeby je tento program volně dostupný internetových stránkách [3].

Program Weka nabízí tři možné programové rozhraní pro návrh modelu, ve kterých je možné pracovat, a to Explorer, Experimenter a KnowledgeFlow. Vzhledem k příjemnému grafickému rozhraní bylo pro návrh modelu zvoleno prostředí KnowledgeFlow.

5.2 Import dat

Dříve než dojde k samotnému návrhu modelu v programovém prostředí Weka, je nezbytnou nutností úprava dat do správného formátu, aby bylo možné data importovat do programu. Budou sestaveny tři soubory, z nichž každý bude představovat jednu ze tří oblastí dat, které byly blíže popsány v kapitole 3.

Program Weka pracuje se soubory ve formátu *arff*. Formátu *arff* je možno dosáhnout pomocí uložení datové matice ve formátu *CSV* (oddělený středníkem). Formát *CSV* je potřeba dále otevřít v textovém editoru a doplnit o záhlaví datové matice.

Záhlaví všech tří datových matic bylo upraveno do formátu, který je vyobrazen na obrázku 5.1. Jednotlivé parametry obsahují hodnoty reálných čísel, proto jsou tyto

atributy definovány hodnotou *real*. Po nadefinování všech parametrů je třeba doplnit záhlaví o speciální parametr *class*, který představuje zařazení objektu do stanovené třídy, a nabývá hodnot od jedné do sedmi. Po dokončení nastavení záhlaví je potřebné již pouze uložit změny ve formátu *arff*.

```
@RELATION obce
@ATTRIBUTE x1 real
@ATTRIBUTE x2 real
@ATTRIBUTE x3 real
@ATTRIBUTE x4 real
@ATTRIBUTE x5 real
@ATTRIBUTE x6 real
@ATTRIBUTE x7 real
@ATTRIBUTE x8 real
@ATTRIBUTE x9 real
@ATTRIBUTE x10 real
@ATTRIBUTE x11 real
@ATTRIBUTE class {1,2,3,4,5,6,7}
@data
```

Obrázek 5.1: Nastavení záhlaví pro import dat. Zdroj: [vlastní]

5.3 Návrh modelu v programovém prostředí Weka KnowledgeFlow

Návrh modelu v programovém prostředí Weka KnowledgeFlow, bude sestaven na základě navrženého modelu v kapitole 4. Pomocí uzlů, které je možné použít na pracovní ploše, byl sestaven návrh modelu, který byl zobrazen v podkapitole 4.6.

Podstatnou roli v návrhu modelu hraje rozdělení vstupní množiny dat na množinu testovací a trénovací. Pomocí trénovacích dat se algoritmu SMO učí, na testovacích datech poté testuje. V případě všech navržených modelů bylo využito procentuální rozdělení, ve kterém testovací množinu dat tvoří 25 % dat, zbylou část 75 % poté množina dat trénovací. Toto rozdělení bylo zvoleno na základě předchozích pokusů nastavení, jelikož udávalo nejlepší výsledky.

Nejdůležitější krok při sestavování návrhu modelu představuje implementace algoritmu SMO, který byl blíže popsán v podkapitole 4.5. Důležité je nastavení parametrů, které ovlivňují výsledky klasifikace.

Algoritmus SMO patří mezi učící se funkce a důležitým krokem při jeho implementaci je nastavení některých parametrů, který mohou výrazně ovlivnit kvalitu dosažených výsledků. Výčet nejdůležitějších parametrů algoritmu SMO, které nejvíce ovlivňují navržený model, je zobrazen v tabulce 5.1.

Tabulka 5.1: Důležité parametry algoritmu SMO Zdroj: [vlastní]

<i>Důležité parametry algoritmu SMO</i>		
C	The complexity constant	nabývá hodnot $(-\infty ; +\infty)$
N (filter Type)	filter Type	nabývá hodnot $\{0, 1, 2\}$ 0 = data beze změny 1 = normalizovaná data 2 = standardizovaná data
L	The tolerance parameter	nabývá hodnot $(-\infty ; +\infty)$
useRBF	Radial Basic Function	nabývá hodnot (True, False)

5.4 Návrh parametrů modelu pro ekonomická data

Při návrhu modelu s Polynomiální jádrovou funkcí byl nastaven parametr *useRBF* u algoritmu SMO na hodnotu *False*, což znamená, že parametr Radial Basic Function je zakázán a nastaven na hodnotu default. Pro dosažení co nejlepších výsledků byly následně pro jednotlivé navržené struktury voleny zbylé důležité parametry, které mohou velmi ovlivnit kvalitu navrženého modelu. Kvalita navržených modelů je vyjádřena procentuální úspěšností testovacích dat. V tabulce 5.2 je zobrazeno šest nejlepších výsledků, kterých bylo dosaženo při nastavení parametru *useRBF* na hodnotu *False*.

Tabulka 5.2: Návrh parametrů modelu pro ekonomická data I. Zdroj: [vlastní]

<i>Struktura</i>	<i>C</i>	<i>N</i>	<i>L</i>	<i>useRBF</i>	<i>Zařazené případy [v %]</i>	
					<i>Správné</i>	<i>Nesprávné</i>
1	1	0	0.0015	false	85.8407	14.1593
2	2	0	0.0010	false	84.9558	15.0442
3	4	2	0.0025	false	85.8407	14.1593
4	8	2	0.0030	false	84.0708	15.9292
5	10	1	0.0015	false	86.7257	13.2743
6	11	1	0,0050	false	87.6106	12.3894

Při návrhu modelu s RBF jádrovou funkcí je zapotřebí nastavení parametru *useRBF* u algoritmu SMO na hodnotu *True*. Dalším krokem je nastavení ostatních parametrů. Následující tabulka 5.3 prezentuje šest nejlepších dosažených výsledků.

Tabulka 5.3: Návrh parametrů modelu pro ekonomická data II. Zdroj: [vlastní]

<i>Struktura</i>	<i>C</i>	<i>N</i>	<i>L</i>	<i>useRBF</i>	<i>Zařazené případy [v %]</i>	
					<i>Správné</i>	<i>Nesprávné</i>
7	2	2	0.0015	True	87.6106	12.3894
8	3	2	0.0020	True	86.7257	13.2743
9	6	0	0.0035	true	89.3807	10.6195
10	8	2	0.0025	true	90.2655	9.7345
11	10	2	0.0015	true	91.1504	8.8496
12	15	0	0,0045	true	88.4956	11.5044

5.4.1 Optimální struktura navrženého modelu pro ekonomická data

Při výběru optimální struktury navrženého modelu je nutností sledovat ukazatele procentuální úspěšnosti klasifikace. Logicky lze odvodit, že za nejlépe navržený model lze považovat takový model, jehož procentuální úspěšnost dosahuje nejvyšší hodnoty. Při pohledu na tabulky 5.2 a 5.3 je patrné, že nejvhodnější návrh parametrů modelu představuje struktura 11.

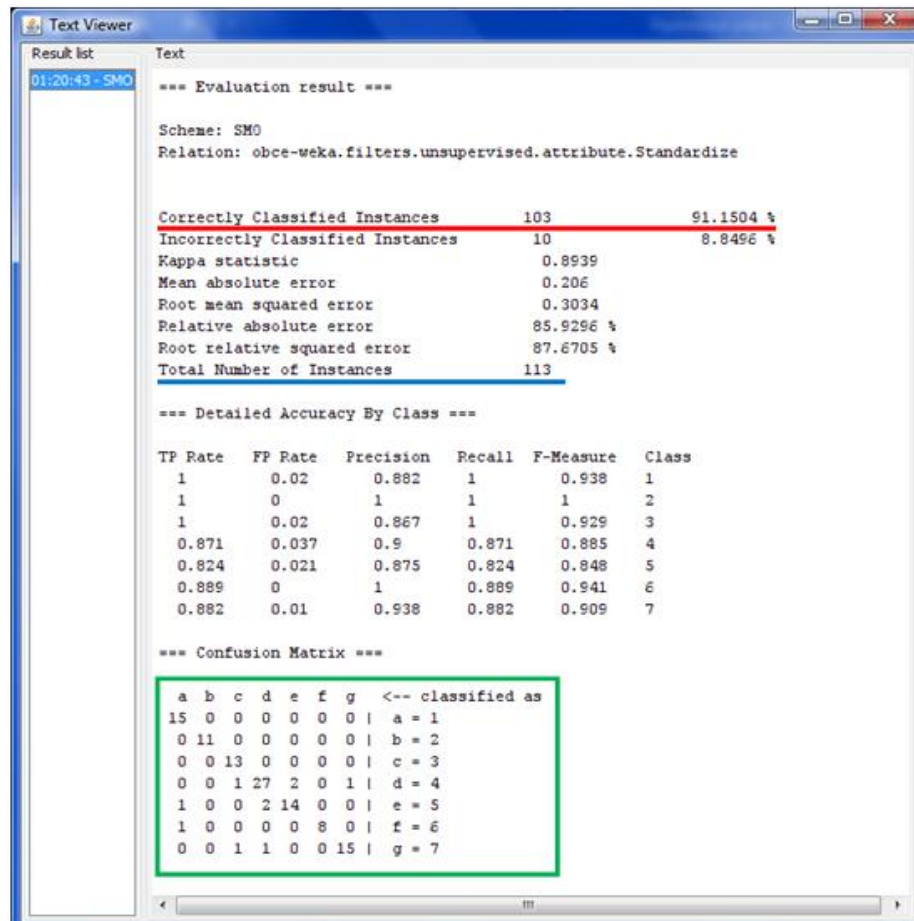
Tabulka 5.4: Optimální struktura modelu klasifikace pro ekonomická data. Zdroj: [vlastní]

<i>Struktura</i>	<i>C</i>	<i>N</i>	<i>L</i>	<i>useRBF</i>	<i>Zařazené případy [v %]</i>	
					<i>Správné</i>	<i>Nesprávné</i>
11	10	2	0.0015	true	91.1504	8.8496

Tabulka 5.4 prezentuje výsledek optimální struktury modelu klasifikace pro ekonomická data udržitelného rozvoje. Při nastavení parametru *useRBF* na hodnotu *True*, nastavení parametrů $C = 10$ a $L = 0.0015$ a vstupních standardizovaných datech, bylo dosaženo procentuální úspěšnosti správně zařazených případů ve výši 91.1504 %.

5.4.2 Analýza klasifikátoru zvoleného modelu pro ekonomická data

Výstup navrženého modelu představuje klasifikátor parametrů ekonomické oblasti udržitelného rozvoje. Na obr. 5.3 je vyobrazena textová podoba výsledků klasifikace při sestavení optimální struktury modelu pro klasifikaci ekonomických dat udržitelného rozvoje.



```
Text Viewer
Result list
01:20:43 - SMO
Text
=== Evaluation result ===
Scheme: SMO
Relation: obce-weka.filters.unsupervised.attribute.Standardize

Correctly Classified Instances      103      91.1504 %
Incorrectly Classified Instances    10       8.8496 %
Kappa statistic                    0.8939
Mean absolute error                 0.206
Root mean squared error             0.3034
Relative absolute error             85.9296 %
Root relative squared error         87.6705 %
Total Number of Instances          113

=== Detailed Accuracy By Class ===

TP Rate  FP Rate  Precision  Recall  F-Measure  Class
1         0.02     0.882     1       0.938     1
1         0        1         1       1         2
1         0.02     0.867     1       0.929     3
0.871    0.037    0.9       0.871   0.885     4
0.824    0.021    0.875    0.824   0.848     5
0.889    0        1         0.889   0.941     6
0.882    0.01     0.938    0.882   0.909     7

=== Confusion Matrix ===
 a b c d e f g <-- classified as
15 0 0 0 0 0 0 | a = 1
0 11 0 0 0 0 0 | b = 2
0 0 13 0 0 0 0 | c = 3
0 0 1 27 2 0 1 | d = 4
1 0 0 2 14 0 0 | e = 5
1 0 0 0 0 8 0 | f = 6
0 0 1 1 0 0 15 | g = 7
```

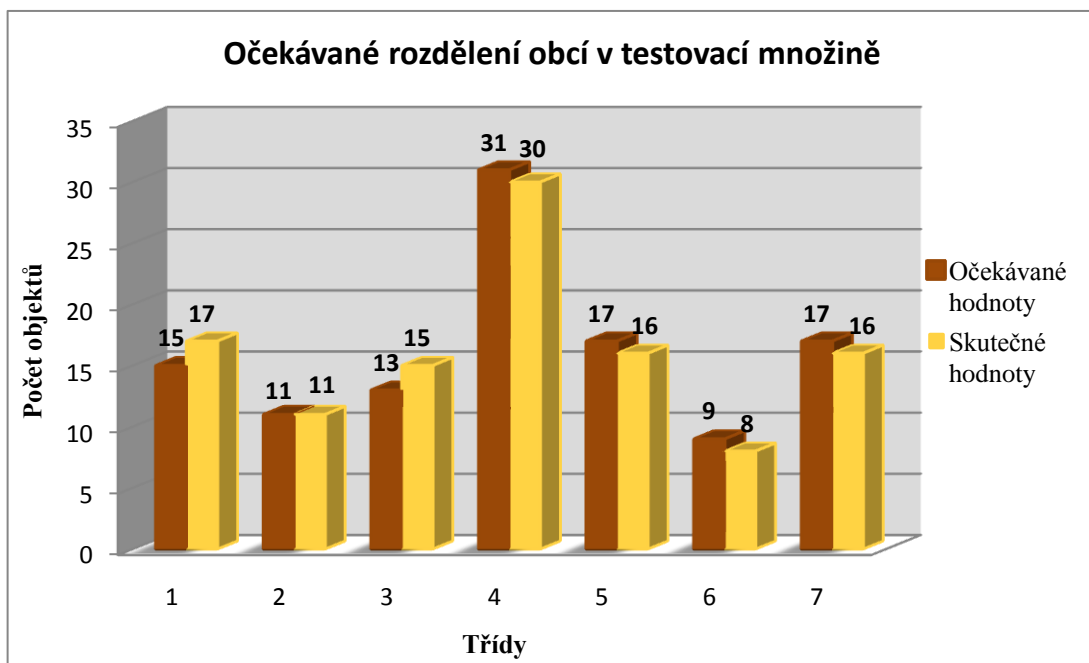
Obrázek 5.2: Výsledek klasifikace ekonomických dat. Zdroj: [vlastní]

Zaměříme-li se na barevně zvýrazněné údaje, je možné o výsledcích říci, že při testovací množině o celkovém počtu 113 objektů, bylo dosaženo úspěšnosti klasifikace ve výši 91.1504%, což odpovídá správné klasifikaci celkem 103 obcí. Zbýlých 10 obcí bylo klasifikováno nekorektně. Důležitou informací, kterou také výsledek poskytuje, je matice klasifikátoru, která nám udává celkový počet objektů v jednotlivých třídách a jejich správnost zařazení. Pro názornost byla matice klasifikátoru transformována do tabulky 5.5, kde řádky jsou tvořeny počty obcí, jež byly dle očekávání zařazeny do jednotlivých tříd. Naproti tomu sloupce vyjadřují roztřídění tříd testovací množiny na základě navrženého klasifikátoru.

Tabulka 5.5: Porovnání očekávaného a skutečného rozložení eko dat. Zdroj: [vlastní]

		Skutečné rozložení [číslo]						
		1	2	3	4	5	6	7
Očekávané rozložení [číslo]	1	15	0	0	0	0	0	0
	2	0	11	0	0	0	0	0
	3	0	0	13	0	0	0	0
	4	0	0	1	27	2	0	1
	5	1	0	0	2	14	0	0
	6	1	0	0	0	0	8	0
	7	0	0	1	1	0	0	15

K tabulce 5.5 byl sestaven graf 5.1, jež vykresluje zjištěné počty objektů testovací skupiny v rámci očekávaného rozložení a skutečného rozložení dle klasifikátoru. Jak je patrné z grafu, v průběhu klasifikace nedošlo k větším nepřesnostem. Za zmínění by v tomto případě stály pouze dvě situace. První je vázána ke třídě 1, do které byly během klasifikace přidány navíc dva objekty, a to po jednom objektu ze tříd 5 a 6. Druhým případem je poté třída 3, kde došlo také během klasifikace k přidání 2 objektů, a to ze 4 a 7 třídy.



Graf 5.1: Očekávané a skutečné rozdělení ekonomických dat. Zdroj: [vlastní]

5.5 Návrh parametrů modelu pro environmentální data

Návrh parametrů modelu klasifikátoru pro environmentální data udržitelného rozvoje byl sestavován na stejné bázi jako v případě ekonomických dat. Při návrhu modelu s Polynomiální jádrovou funkcí byl nastaven parametr *useRBF* u algoritmu SMO na hodnotu *False* a následně byly voleny také ostatní parametry, které ovlivňují kvalitu navrženého modelu.

V tabulce 5.6 je zobrazeno šest nejlepších výsledků, kterých bylo dosaženo při nastavení parametru *useRBF* na hodnotu *False*.

Tabulka 5.6: Návrh parametrů modelu pro environmentální data I. Zdroj: [vlastní]

Struktura	C	N	L	useRBF	Zařazené případy [v %]	
					Správné	Nesprávné
1	2	0	0.0015	false	90.2655	9.7345
2	3	0	0.0045	false	89.3805	10.6195
3	1	1	0.0020	false	84.0708	15.9292
4	3	1	0.0010	false	90.2655	9.7345
5	7	2	0.0030	false	91.1504	8.8496
6	38	0	0,0050	false	92.0354	7.9646

Při změně nastavení parametru *useRBF* u algoritmu SMO na hodnotu *True* je stejně jako u ekonomických dat zapotřebí opět vyzkoušet různá nastavení parametrů s cílem dosažení nejvhodnější struktury. Šest nejlepších výsledků je znázorněno v tabulce 5.7.

Tabulka 5.7: Návrh parametrů modelu pro environmentální data II. Zdroj: [vlastní]

Struktura	C	N	L	useRBF	Zařazené případy [v %]	
					Správné	Nesprávné
7	16	0	0.0030	True	92.9204	7.0796
8	40	1	0.0020	True	90.2655	9.7345
9	55	1	0.0035	true	91.1504	8.8496
10	16	2	0.0045	true	92.0354	9.7345
11	17	2	0.0010	true	92.9204	7.0796
12	21	2	0,0035	true	93.8053	6.1947

5.5.1 Optimální struktura navrženého modelu pro environmentální data

Při výběru optimální struktury navrženého modelu je nutností sledovat ukazatele procentuální úspěšnosti klasifikace. Na základě tabulek 5.6 a 5.7 byl zvolen nejvhodnější návrh klasifikátoru parametrů.

Tabulka 5.8: Optimální struktura modelu klasifikace pro envi data. Zdroj: [vlastní]

Struktura	C	N	L	useRBF	Zařazené případy [v %]	
					Správné	Nesprávné
12	21	2	0.0035	true	93.8053	6.1947

Tabulka 5.8 prezentuje výsledek optimální struktury modelu klasifikace pro environmentální data udržitelného rozvoje. Při nastavení parametru *useRBF* na hodnotu *True*, nastavení parametrů *C* = 21 a *L* = 0.0035 a vstupních standardizovaných datech, bylo dosaženo procentuální úspěšnosti správně zařazených případů ve výši 93.8053%.

5.5.2 Analýza klasifikátoru zvoleného modelu pro environmentální data

```

Text Viewer
Result list  Text
04:48:31 - SMO  === Evaluation result ===
04:49:02 - SMO
04:49:25 - SMO  Scheme: SMO
05:24:10 - SMO  Relation: obce2-weka.filters.unsupervised.attribute.Standardize
05:25:00 - SMO
05:25:17 - SMO
05:26:15 - SMO

Correctly Classified Instances      106      93.8053 %
Incorrectly Classified Instances     7        6.1947 %
Kappa statistic                     0.924
Mean absolute error                  0.2053
Root mean squared error              0.3023
Relative absolute error              86.3942 %
Root relative squared error          87.7031 %
Total Number of Instances           113

=== Detailed Accuracy By Class ===

TP Rate  FP Rate  Precision  Recall  F-Measure  Class
1        0.009    0.857     1       0.923     1
1        0        1         1       1         2
0.964    0.012    0.964    0.964   0.964     3
0.929    0.02     0.867    0.929   0.897     4
0.667    0        1         0.667   0.8       5
0.88     0.011    0.957    0.88    0.917     6
1        0.022    0.923     1       0.96     7

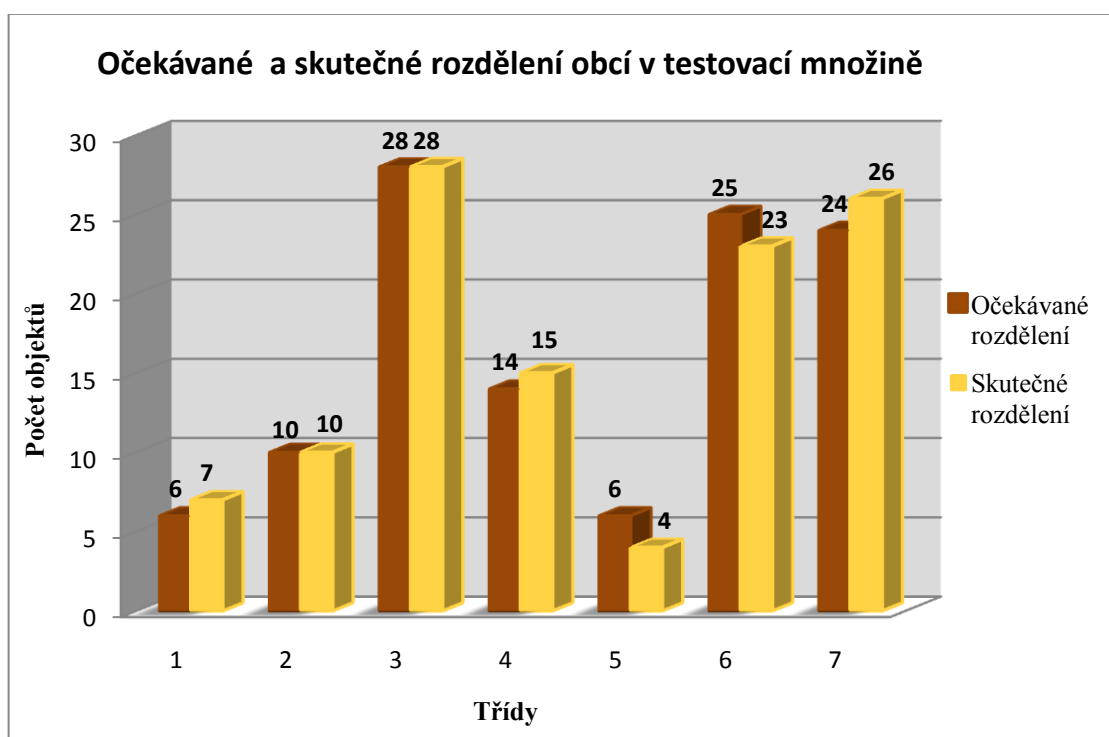
=== Confusion Matrix ===
 a  b  c  d  e  f  g  <-- classified as
ε  0  0  0  0  0  0 | a = 1
0 10  0  0  0  0  0 | b = 2
0  0 27  0  0  0  1 | c = 3
0  0  0 13  0  1  0 | d = 4
0  0  0  1  4  0  1 | e = 5
1  0  1  1  0 22  0 | f = 6
0  0  0  0  0  0 24 | g = 7
    
```

Obrázek 5.3: Výsledek klasifikace environmentálních dat. Zdroj: [vlastní]

Výstup navrženého modelu představuje klasifikátor parametrů environmentální oblasti udržitelného rozvoje. Obrázek 5.4 zobrazuje textovou podobu výsledků klasifikace při sestavení nejhodnější struktury modelu pro klasifikaci.

Dle výsledků je možné říci, že klasifikace testovací množiny byla provedena s úspěšností 93.8053%. Z celkového počtu objektů v testovací množině bylo korektně klasifikováno 106 obcí a nekorektně zbylých 7 obcí.

Pro názornost byla matice klasifikátoru opět transformována do tabulky, viz tabulka 5.9, ke které byl sestaven graf 5.2.



Graf 5.2: Očekávané a skutečné rozdělení environmentálních dat. Zdroj: [vlastní]

Při porovnávání očekávaného a skutečného rozdělení došlo k podobné situaci jako u dat ekonomických. Ve třech situacích došlo ke změně o 2 objekty oproti očekávanému rozdělení. Ve třídě 5 došlo ke snížení počtu obcí o 2 objekty z očekávaných 6 na 4. Ke stejnému snížení došlo také ve třídě 6, kde bylo očekáváno celkem 25 objektů a zařazeno bylo do této třídy pouze 23 objektů. Naopak ve třídě 7 došlo k přidání 2 objektů a celkový počet objektů zařazených do této skupiny byl 26. Ve dvou případech došlo k navýšení objektů ve třídě o 2 objekt a ve dvou případech byl očekávaný počet objektů stejný s počtem skutečným.

Tabulka 5.9: Porovnání očekávaného a skutečného rozložení envi dat. Zdroj: [vlastní]

		Skutečné rozložení [číslo]						
		1	2	3	4	5	6	7
Očekávané rozložení [číslo]	1	6	0	0	0	0	0	0
	2	0	10	0	0	0	0	0
	3	0	0	27	0	0	0	1
	4	0	0	0	13	0	1	0
	5	0	0	0	1	4	0	1
	6	1	0	1	1	0	22	0
	7	0	0	0	0	0	0	24

Pro názornost byla matice klasifikátoru opět transformována do tabulky, viz tabulka 5.9, ke které byl sestaven graf 5.2.

5.6 Návrh parametrů modelu pro sociální data

Návrh parametrů modelu klasifikátoru pro sociální data udržitelného rozvoje, vzhledem k sestavování předchozích skupin dat, je neměnný. Stejným způsobem jako u předchozích skupin byl při návrhu modelu s Polynomiální jádrovou funkcí nastaven parametr *useRBF* u algoritmu SMO na hodnotu *False* a následně byly nastavovány také ostatní parametry, které ovlivňují kvalitu navrženého modelu.

V tabulce 5.10 je zobrazeno šest nejlepších výsledků, kterých bylo dosaženo při nastavení parametru *useRBF* na hodnotu *False*.

Druhým možným případem je nastavení parametru *useRBF* u algoritmu SMO na hodnotu *True*. Šest nejlepších výsledků je znázorněno v tabulce 5.11.

Tabulka 5.10: Návrh parametrů modelu pro sociální data I. Zdroj: [vlastní]

Struktura	C	N	L	useRBF	Zařazené případy [v %]	
					Správné	Nesprávné
1	1	0	0.0040	false	86.7257	13.2743
2	4	0	0.0025	false	84.9558	15.0442
3	9	0	0.0055	false	84.0708	15.9292
4	4	1	0.0035	false	86.7257	13.2743
5	8	1	0.0010	false	85.8407	14.1593
6	1	2	0.0015	false	87.6106	12.3894

Tabulka 5.11: Návrh parametrů modelu pro sociální data II. Zdroj: [vlastní]

Struktura	C	N	L	useRBF	Zařazené případy [v %]	
					Správné	Nesprávné
7	5	2	0.0015	True	86.7257	13.2743
8	19	1	0.0040	True	84.9558	15.0442
9	15	1	0.0025	True	85.8407	14.1593
10	1	0	0.0055	True	88.4956	11.5044
11	2	0	0.0035	True	83.1858	16.8142
12	3	0	0.0010	True	84.9558	15.0442

5.6.1 Optimální struktura navrženého modelu pro sociální data

Optimální strukturou navrženého modelu je na základě procentuální úspěšnosti pro sociální data struktura 10.

Tabulka 5.12: Optimální struktura modelu klasifikace pro sociální data. Zdroj: [vlastní]

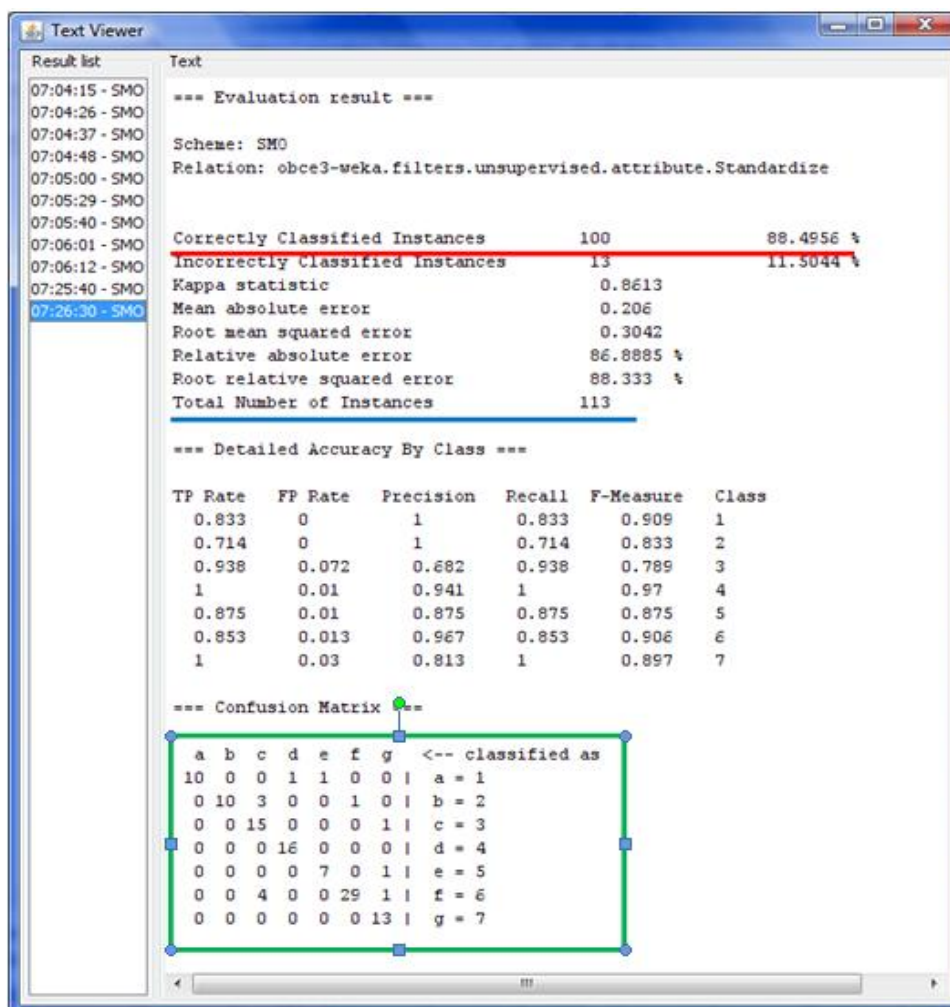
Struktura	C	N	L	useRBF	Zařazené případy [v %]	
					Správné	Nesprávné
10	1	0	0.0055	True	88.4956	11.5044

Tabulka 5.12 prezentuje výsledek optimální struktury modelu klasifikace pro sociální data udržitelného rozvoje, kterým je nastavení parametru *useRBF* na hodnotu *True*, nastavení parametrů $C = 1$ a $L = 0.0055$ a vstupní data nejsou ani standardizovaná ani normalizovaná. V závislosti na vybranou oblast dat bylo dosaženo procentuální úspěšnosti správně zařazených případů ve výši 88.4956%.

5.6.2 Analýza klasifikátoru zvoleného modelu pro sociální data

Výstup navrženého modelu představuje klasifikátor parametrů sociální oblasti udržitelného rozvoje. Na obrázku 5.4 je vyobrazena textová podoba výsledků klasifikace při sestavení nejvhodnější struktury modelu pro klasifikaci.

Dle výsledků klasifikátoru sociálních parametrů udržitelného rozvoje bylo korektně zařazeno 100 obcí z celkové množiny testovacích dat. Zbýlých 13 obcí nebylo zařazeno podle očekávání. Z toho také vyplývá, že úspěšnost klasifikace bude nižší než v předchozích skupinách dat. Nejvhodnější nastavení modelu přineslo úspěšnost klasifikace 88.4956.



```
Text Viewer
Result list
Text
=== Evaluation result ===
Scheme: SMO
Relation: obce3-weka.filters.unsupervised.attribute.Standardize
Correctly Classified Instances      100      88.4956 %
Incorrectly Classified Instances    13       11.5044 %
Kappa statistic                    0.8613
Mean absolute error                 0.206
Root mean squared error             0.3042
Relative absolute error             86.8885 %
Root relative squared error         88.333 %
Total Number of Instances          113
=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.833    0        1          0.833  0.909     1
0.714    0        1          0.714  0.833     2
0.938    0.072    0.682     0.938  0.789     3
1        0.01     0.941     1      0.97      4
0.875    0.01     0.875     0.875  0.875     5
0.853    0.013    0.967     0.853  0.906     6
1        0.03     0.813     1      0.897     7
=== Confusion Matrix ===
 a b c d e f g <-- classified as
10 0 0 1 1 0 0 | a = 1
 0 10 3 0 0 1 0 | b = 2
 0 0 15 0 0 0 1 | c = 3
 0 0 0 16 0 0 0 | d = 4
 0 0 0 0 7 0 1 | e = 5
 0 0 4 0 0 29 1 | f = 6
 0 0 0 0 0 0 13 | g = 7
```

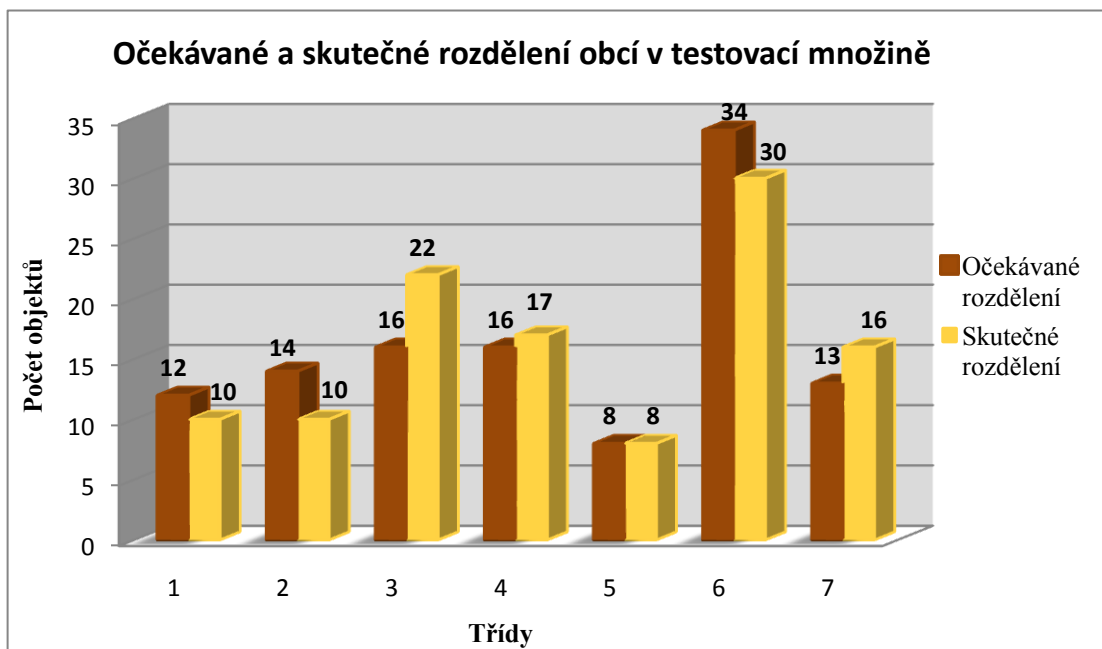
Obrázek 5.4: Výsledek klasifikace sociálních dat. Zdroj: [vlastní]

Přehlednější vyjádření matice klasifikátoru je zobrazeno v tabulce 5.13.

Tabulka 5.13: Porovnání očekávaného a skutečného rozložení soc dat. Zdroj: [vlastní]

		Skutečné rozložení [číslo]						
		1	2	3	4	5	6	7
Očekávané rozložení [číslo]	1	10	0	0	1	1	0	0
	2	0	10	3	0	0	1	0
	3	0	0	15	0	0	0	1
	4	0	0	0	16	0	0	0
	5	0	0	0	0	7	0	1
	6	0	0	4	0	0	29	1
	7	0	0	0	0	0	0	13

Při pohledu na graf 5.3 je zřejmé, že v případě ekonomických dat se jedná o jinou situaci než v předchozích dvou skupinách. Při porovnávání očekávaného a skutečného rozdělení došlo v některých třídách k velikým změnám. Nejvíce se průběh klasifikace dotknul třídy 3, do které bylo zařazeno o 6 obcí více, než se očekávalo. Naproti tomu ve třídách 2 a 6 došlo ke snížení počtu obcí o 4 objekty.



Graf 5.3: Očekávané a skutečné rozdělení sociálních dat. Zdroj: [vlastní]

5.7 Dílčí závěr

Tato kapitola byla věnována návrhu nejvhodnějšího modelu klasifikátoru parametrů udržitelného rozvoje. Jelikož data nebyla posuzována jako jednotný celek, ale bylo na ně pohlíženo pouze v rámci oblasti, do které spadají, bylo potřeba vytvořit tři modely. Ke každému souboru dat byl vybrán nejvhodnější model, jehož úspěšnost klasifikace byla nejvyšší. V rámci ekonomických dat bylo dosaženo úspěšnosti klasifikace 91.504% a u environmentálních dat bylo dosaženo 93.8053%. Úspěšnosti klasifikace v rámci sociálních dat byla 88.4956%.

6 Aproximace navržených modelů

Tato kapitola je zaměřena na aproximaci dat navržených modelů. V rámci aproximace bylo cílem zjistit, zda parametry získané od experta byly navrženy pro klasifikaci správně. V rámci jednotlivých podkapitol budou vytvářena aproximace ekonomických, environmentálních i sociálních dat.

Vstupní data jsou vyjádřena vstupní maticí, která je tvořena daným počtem obcí v Pardubickém kraji. Na základě rozhodnutí experta bylo vybráno celkem 42 parametrů, z nichž 11 parametrů bylo z oblasti ekonomické, 11 z oblasti environmentální a 20 z oblasti sociální. Tato problematika je blíže specifikována v kapitole 3.

6.1 Korelace

Datová matice, která je složena z daných obcí a stanovených parametrů, byla podkladem pro následné zjištění míry závislosti navržených parametrů. Jelikož na data pohlížíme ve třech různých oblastech, byly sestaveny míry závislostí pro každou oblast zvlášť. K vyjádření míry závislosti je používán korelační koeficient. Korelační koeficient dvou náhodných veličin X , Y je vyjádřen vztahem

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{DX}\sqrt{DY}}, \quad (6.1)$$

kde $\rho_{X,Y} = 0$, náhodné veličiny X a Y jsou nekorelované, v opačném případě mezi těmito veličinami existuje korelační vztah [13].

Vypočítané korelace pro jednotlivé modely jsou v příloze 2. Na základě doporučení experta byly použity všechny parametry, přestože mezi některými parametry byly velmi silné korelační vztahy. Korelační matice budou dále použity při rozhodování v situacích, při kterých je potřeba rozhodovat o vyloučení mezi dvěma či více parametry.

6.2 Analýza aproximace

Cílem analýzy je získání informace, jak bude ovlivněna přesnost klasifikace v případě, že bude vyřazen jeden ze vstupních parametrů, které byly navrženy expertem. Postupně z datové matice byl odebrán vždy pouze jeden ze vstupních parametrů a byla zjištěna přesnost klasifikace testovacích dat. Následně je tento výsledek porovnán s původním nastavením, kdy byly načteny všechny vstupní parametry. Jednotlivé výsledky jsou následně zpracovány do jednoduché tabulky, viz tabulka 6.1.

Tabulka 6.1: Ukázkový příklad prezentace výsledků analýzy. Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchylka [v %]
VK	3	1	0.0010	true	90,2436	ZH
VK-1	3	1	0.0010	true	80,7869	-9,4567
VK-2	3	1	0.0010	true	91,5646	1,3210
VK-3	3	1	0.0010	true	90,2436	0,0000

První sloupec tabulky vyjadřuje počet vstupních parametrů. VP znázorňuje označení pro všechny parametry. Následující čtyři sloupce vyjadřují optimální nastavení algoritmu SMO. Předposlední sloupec vykresluje výsledné přesnosti klasifikace testovacích dat a výsledky jsou uváděny v procentech. Poslední sloupec vyjadřuje odchylku. V případě načtení na počátku všech vstupních parametrů bez vyřazení některého z parametrů je brána přesnost klasifikace VK jako základní hodnota. Ve chvíli kdy dojde k odstranění některého z parametrů, dochází ke zjišťování odchylky přesnosti klasifikace. Tato odchylka je udávána také v procentech a vyjadřuje, o kolik procent se zvýšila či snížila přesnost klasifikace v případě, že došlo k odstranění daného parametru. Tato hodnota může být vyjádřena:

- **v záporných hodnotách** – hodnoty odchylky jsou vyjádřeny znaménkem mínus a představují hodnoty menší než 0, což znamená, že odstraněním daného parametru došlo ke snížení přesnosti klasifikace, čímž výsledek vede k závěru, že daný parametr má vliv na zvýšení přesnosti klasifikace a není doporučeno jej vyřazovat;

- **v kladných hodnotách** - tyto hodnoty představují odchylky větší než 0, což znamená, že odstraněním daného parametru došlo ke zvýšení přesnosti klasifikace, čímž výsledek vede k závěru, že odstranění parametru má vliv na zvýšení přesnosti klasifikace a doporučuje se jej vyřadit;
- **nulovou hodnotu** - tyto hodnoty představují odchylky rovnou 0 a znamená to, že odstraněním daného parametru došlo k získání stejné přesnosti klasifikace jako v případě, kdy jsou zavedeny všechny VK, což vede k závěru, že odstranění parametru nemá vliv na přesnost klasifikace a doporučuje se jej vyřadit.

V případě, že by došlo k vyřazení některého z parametrů, postupuje se stejným způsobem opět znovu ovšem s malým rozdílem. V další části by docházelo k posuzování přesností klasifikace po vyloučení druhého parametrů s přesností, která byla získána při odstranění jednoho parametru. V případě, že by došlo ke stejným výsledkům odchylek u více jak jednoho parametru, bude se přihlížet ke korelacím dat, které jsou zobrazeny v příloze 2.

6.3 Aproximace ekonomického modelu

V rámci aproximace ekonomického modelu byla provedena analýza, při které bylo zjišťováno, které ze zadaných parametrů ovlivňují přesnost klasifikace. K této analýze bylo použito optimální nastavení algoritmu SMO, na jehož základě byly získány nejlepší výsledky přesnosti klasifikace testovacích dat. Bližší informace o nastavení modelu klasifikátoru pro ekonomická data jsou znázorněny v podkapitole 5.4.

Optimální nastavení algoritmu SMO pro ekonomický model bylo nastaveno na následující hodnoty:

- $C = 10$
- $N = 2$
- $L = 0.0015$
- `useRBF = true`

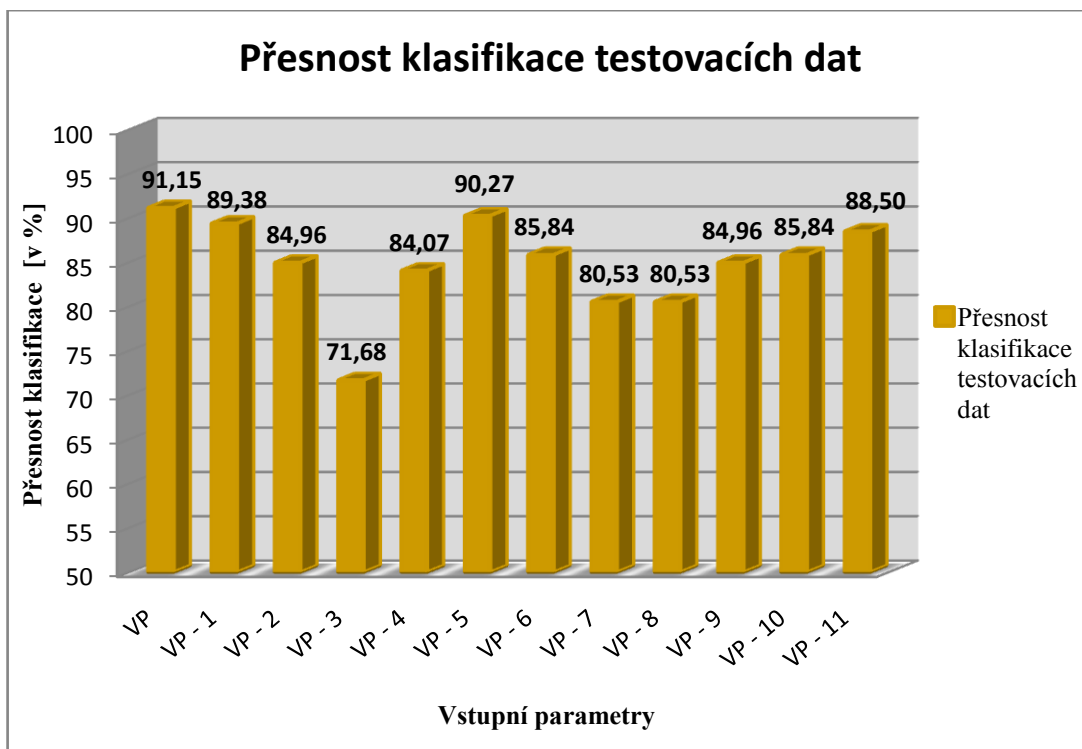
Výsledky přesnosti klasifikace testovacích dat v rámci ekonomického modelu jsou prezentovány tabulkou 6.2. Základní hodnotu v tomto případě vyjadřuje výsledek přesnosti klasifikace testovacích dat, kterého bylo dosaženo na základě využití všech VP, které byly navrženy expertem. V počáteční fázi bylo načteno všech 11 parametrů a zjištěna přesnost klasifikace při optimálním nastavení. Dále byl postupně odstraněn vždy jeden parametr, zjištěna přesnost klasifikace a odchylka.

Odchylky přesnosti se pohybují v záporných hodnotách v intervalu $\langle -19,4690; -0,8849 \rangle$, jak je možné vidět v tabulce 6.2. Z této skutečnosti lze říci, že vyloučením jakéhokoli parametru by bylo docíleno pouze snížení přesnosti klasifikace a návrh experta byl správný.

Tabulka 6.2: Přesnosti klasifikace ekonomických testovacích dat. Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchylka [v %]
VP	10	2	0.0015	True	91,1504	ZH
VP - 1	10	2	0.0015	True	89,3805	-1,7699
VP - 2	10	2	0.0015	True	84,9558	-6,1946
VP - 3	10	2	0.0015	True	71,6814	-19,469
VP - 4	10	2	0.0015	True	84,0708	-7,0796
VP - 5	10	2	0.0015	True	90,2655	-0,8849
VP - 6	10	2	0.0015	True	85,8407	-5,3097
VP - 7	10	2	0.0015	True	80,5310	-10,6194
VP - 8	10	2	0.0015	True	80,5310	-10,6194
VP - 9	10	2	0.0015	True	84,9558	-6,1946
VP - 10	10	2	0.0015	True	85,8407	-5,3097
VP - 11	10	2	0.0015	True	88,4956	-2,6548

Pro porovnání výsledků přesností klasifikací byl sestaven graf 6.1, který vyjadřuje vizuální zobrazení jednotlivých výsledků. Z grafu je možné vidět, že nejlepší přesnosti klasifikace bylo dosaženo při zachování všech parametrů navržených expertem. Naopak ke zhoršení přesnosti klasifikace o více jak 19 % by došlo v případě navržení na vyřazení parametrů p_4 .



Graf 6.1: Vizualizace přesností klasifikace ekonomických testovacích dat. Zdroj: [vlastní]

6.4 Aproximace environmentálního modelu

Aproximace environmentálního modelu byla provedena stejným způsobem jako v případě aproximace modelu ekonomického. Také v tomto případě se vycházelo z optimálního nastavení algoritmu SMO, které je blíže specifikováno v podkapitole 5.5. Optimální nastavení algoritmu SMO pro environmentální model bylo nastaveno na následující hodnoty:

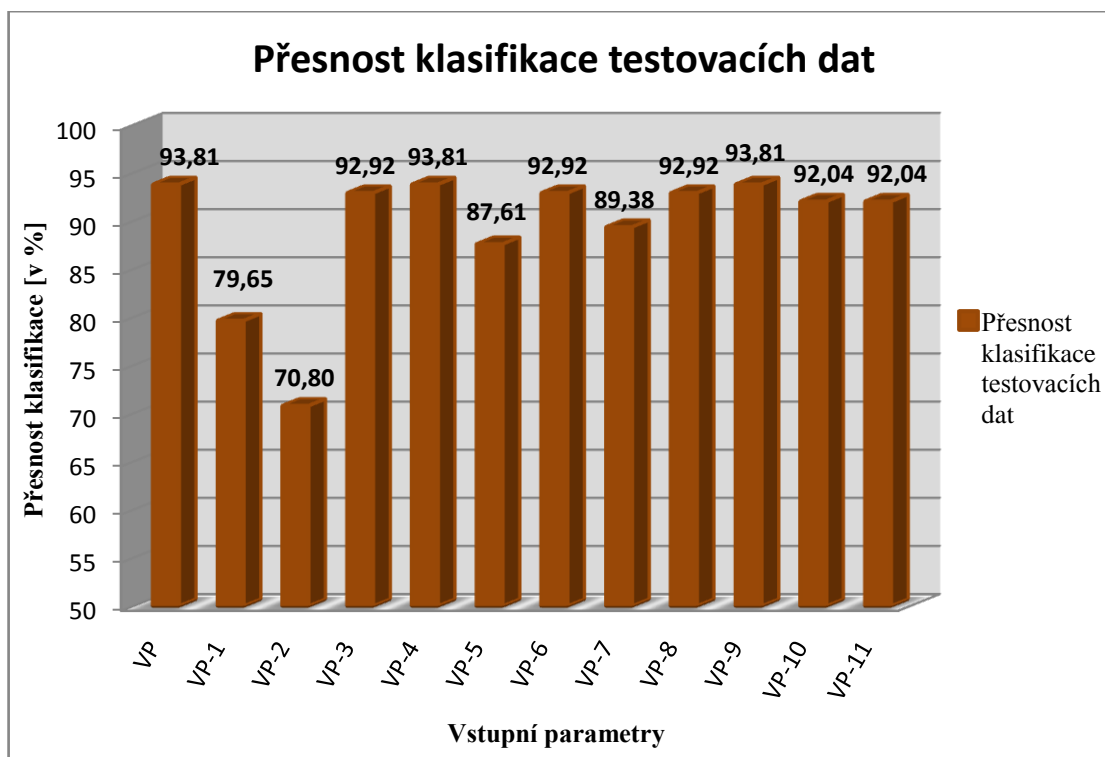
- $C = 21$
- $N = 2$
- $L = 0.0035$
- $useRBF = true$

Výsledky přesnosti klasifikace testovacích dat v rámci environmentálního modelu jsou prezentovány tabulkou 6.3. Environmentální model je tvořen 11 parametry navrženými expertem, které jsou blíže specifikovány v podkapitole 3.2.2. Základní hodnota je tvořena přesností klasifikace, které bylo dosaženo na základě využití všech VK.

Tabulka 6.3: Přesnosti klasifikace environmentálních testovacích dat. Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchyłka [v %]
VP	21	2	0.0035	true	93,8053	ZH
VP – 1	21	2	0.0035	true	79,6460	-14,1593
VP – 2	21	2	0.0035	true	70,7965	-23,0088
VP – 3	21	2	0.0035	true	92,9204	-0,8849
VP – 4	21	2	0.0035	true	93,8053	0,0000
VP – 5	21	2	0.0035	true	87,6106	-6,1947
VP – 6	21	2	0.0035	true	92,9204	-0,8849
VP – 7	21	2	0.0035	true	89,3805	-4,4248
VP - 8	21	2	0.0035	true	92,9204	-0,8849
VP - 9	21	2	0.0035	true	93,8053	0,0000
VP - 10	21	2	0.0035	true	92,0354	-1,7699
VP - 11	21	2	0.0035	true	92,0354	-1,7699

Pro porovnání výsledků přesnosti klasifikací byl sestaven graf 6.2, který vyjadřuje vizuální zobrazení výsledků.



Graf 6.2: Vizualizace přesností klasifikace environmentálních testovacích dat. Zdroj: [vlastní]

Odchytky přesnosti se pohybují v záporných a nulových hodnotách v intervalu $(-23,0088; 0,0000)$, jak je možné vidět v tabulce 6.3. V případě záporných odchylek, lze říci, že vyloučení těchto parametrů by způsobilo pouze snížení přesnosti klasifikace, proto tyto parametry v této fázi nebudou vylučovány. Dalším krokem je zaměřit se na parametry, u kterých po vyřazení došlo k nulové odchylce, a přesnost klasifikace se nezměnila. V tomto případě jde o čtvrtý a devátý parametr, které odpovídají parametrům p_{15} a p_{20} . Pro vyloučení byl navržen čtvrtý neboli p_{15} .

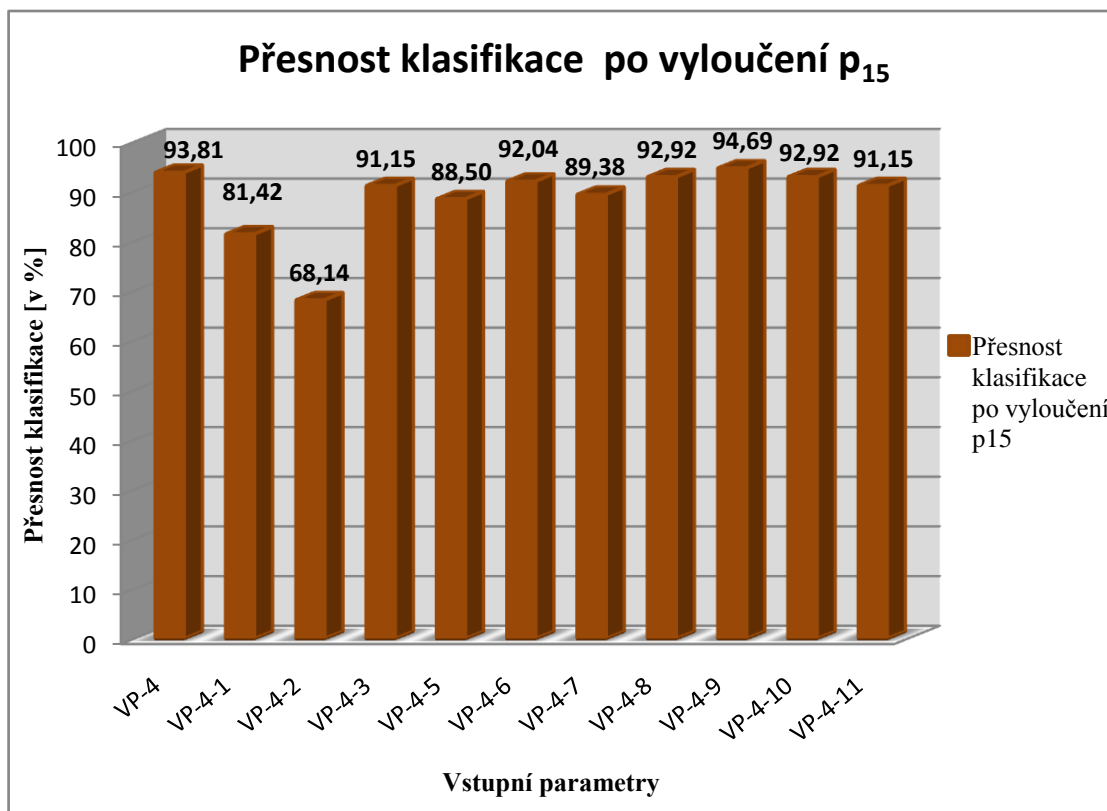
Po návrhu vyřazení parametru p_{15} byl následující postup prováděn ve stejných krocích. Základní hodnota byla tentokrát vyjádřena jako $VK - 4$, což představovalo přesnost klasifikace zbylých deseti parametrů po vyloučení v pořadí čtvrtého parametru. Následně bylo zkoumáno, jak ovlivní přesnost klasifikace, bude-li vyřazen další parametr. Opět byly zjištěny přesnosti klasifikací a odchytky přesností. Výsledky prezentuje tabulka 6.4.

Tabulka 6.4: Přesnosti klasifikace po vyloučení parametru p_{15} . Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchytka [v %]
VP-4	21	2	0.0035	true	93,8053	ZH
VP-4-1	21	2	0.0035	true	81,4159	-12,3894
VP-4-2	21	2	0.0035	true	68,1416	-25,6637
VP-4-3	21	2	0.0035	true	91,1504	-2,6549
VP-4-5	21	2	0.0035	true	88,4956	-5,3097
VP-4-6	21	2	0.0035	true	92,0354	-1,7699
VP-4-7	21	2	0.0035	true	89,3805	-4,4248
VP-4-8	21	2	0.0035	true	92,9204	-0,8849
VP-4-9	21	2	0.0035	true	94,6903	0,8850
VP-4-10	21	2	0.0035	true	92,9204	-0,8849
VP-4-11	21	2	0.0035	true	91,1504	-2,6549

Odchytky přesnosti při pokusu vyloučení druhého parametru se pohybují v hodnotách v intervalu $(-25,6637; 0,8850)$, jak je možné vidět v tabulce 6.4. V případě vyloučení devátého parametru neboli p_{20} došlo jako v jediném případě ke zlepšení přesnosti klasifikace. Na základě této skutečnosti bylo rozhodnuto navrhnout vyřazení tohoto parametru, čímž by bylo dosaženo zlepšení klasifikace na 94,6903 %.

Z grafu 6.3, který prezentuje přesnosti klasifikace po vyloučení parametru p_{15} , je možné vidět vizuální vyjádření výsledků přesností klasifikace po vyloučení dvou parametru. Také z tohoto grafického vyjádření je zřejmé, že největší přesnosti bylo dosaženo po vyloučení čtvrtého a devátého parametru, kterým odpovídají parametry p_{15} a p_{20} .



Graf 6.3: Přesnost klasifikace po vyloučení p_{15} . Zdroj: [vlastní]

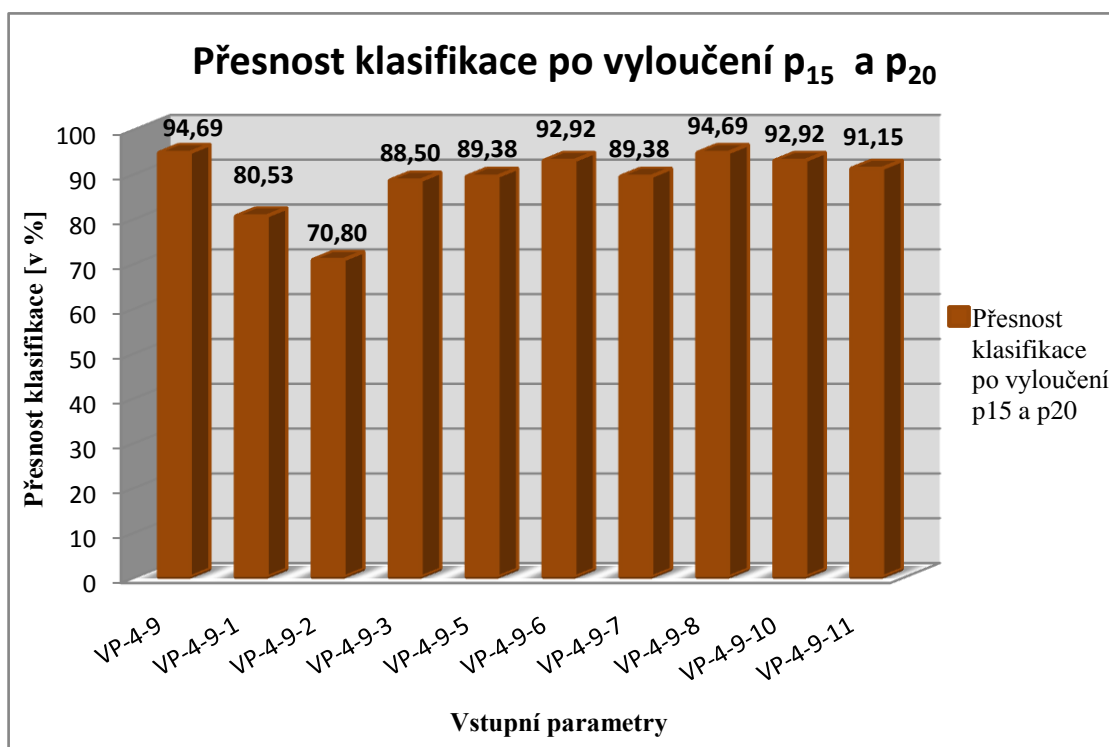
V této fázi byly navrženy již dva parametry k vyřazení a je potřeba provést další analýzu pro vyřazení dalšího parametru. Jako základní hodnotu je brán výsledek předchozí analýzy vyjádřený jako VK – 4 – 9. Prezentaci nových výsledků klasifikace po vyloučení dalšího parametru znázorňuje tabulka 6.5.

Odchytky přesnosti při pokusu vyloučení třetího parametru se pohybují v hodnotách v intervalu $(-23,8938; 0,0000)$, jak je možné vidět v tabulce 6.4. Pouze v jednom případě došlo k získání nulové hodnoty odchytky přesnosti, v ostatních případech byly hodnoty záporné. V případě vyloučení osmého parametru neboli p_{19} nedojde ke změně přesnosti klasifikace, proto je tento parametr navržen jako další na vyřazení.

Tabulka 6.5: Přesnosti klasifikace po vyloučení parametru p_{15} a p_{20} . Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchylka [v %]
VP-4-9	21	2	0.0035	true	94,6903	ZH
VP-4-9-1	21	2	0.0035	true	80,5310	-14,1593
VP-4-9-2	21	2	0.0035	true	70,7965	-23,8938
VP-4-9-3	21	2	0.0035	true	88,4956	-6,1947
VP-4-9-5	21	2	0.0035	true	89,3805	-5,3098
VP-4-9-6	21	2	0.0035	true	92,9204	-1,7699
VP-4-9-7	21	2	0.0035	true	89,3805	-5,3098
VP-4-9-8	21	2	0.0035	true	94,6903	0,0000
VP-4-9-10	21	2	0.0035	true	92,9204	-1,7699
VP-4-9-11	21	2	0.0035	true	91,1504	-3,5399

Graf 6.3 znázorňuje přesnost klasifikace po vyloučení p_{15} a p_{20} . Při pohledu na graf je vidět, že pouze v jednom případě bylo dosaženo stejné klasifikace jako u základní hodnoty, a parametr je jediným adeptem na vyřazení.



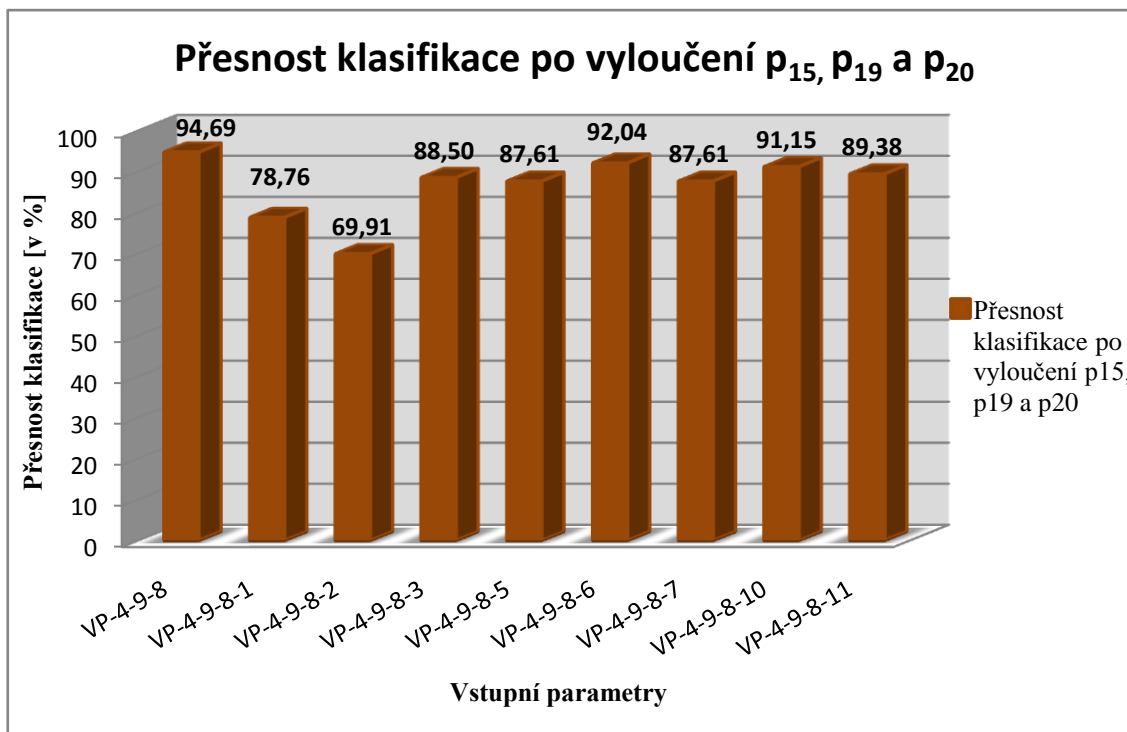
Graf 6.4: Přesnost klasifikace po vyloučení p_{15} a p_{20} . Zdroj: [vlastní]

Po třetí části analýzy byla základní hodnota vyjádřena jako VK – 4 – 9 – 8. Při návrhu vyřazení dalšího parametru byl postup stejný jako v předchozích případech. Výsledky zobrazuje tabulka 6.6.

Tabulka 6.6: Přesnosti klasifikace po vyloučení parametru p_{15} , p_{19} a p_{20} . Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchylna [v %]
VP-4-9-8	21	2	0.0035	true	94,6903	ZH
VP-4-9-8-1	21	2	0.0035	true	78,7611	-15,9292
VP-4-9-8-2	21	2	0.0035	true	69,9115	-24,7788
VP-4-9-8-3	21	2	0.0035	true	88,4956	-6,1947
VP-4-9-8-5	21	2	0.0035	true	87,6106	-7,0797
VP-4-9-8-6	21	2	0.0035	true	92,0354	-2,6549
VP-4-9-8-7	21	2	0.0035	true	87,6106	-7,0797
VP-4-9-8-10	21	2	0.0035	true	91,1504	-3,5399
VP-4-9-8-11	21	2	0.0035	true	89,3805	-5,3098

Grafické znázornění posledních výsledků přesností klasifikace vykresluje graf 6.5.



Graf 6.5: Přesnost klasifikace po vyloučení p_{15} , p_{19} a p_{20} . Zdroj: [vlastní]

Interval $\langle -24,7788; -2,6549 \rangle$ představuje odchylky přesností klasifikací při vyloučení čtyř parametrů. Všechny odchylky se pohybují v záporných hodnotách, což znamená, že vyloučení jakéhokoliv dalšího parametru by vedl pouze ke snížení přesnosti klasifikace, proto není vhodné navrhnout vyřazení některého z dalších parametrů.

Z celé analýzy vyplývá, že nejlepší přesnosti klasifikace pro testovací data bylo dosaženo při odebrání parametrů **p₁₅**, **p₁₉** a **p₂₀**. Z tohoto důvodu byl podán návrh na vyřazení těchto tří parametrů z původní skupiny 11 parametrů.

6.5 Aproximace sociálního modelu

Aproximace sociálního modelu byla provedena stejným způsobem jako v předchozích případech. Optimální nastavení algoritmu SMO pro environmentální model bylo blíže specifikováno v podkapitole 5.6. Bylo nastaveno na následující hodnoty:

- $C = 1$
- $N = 0$
- $L = 0.0055$
- $useRBF = true$

Výsledky přesnosti klasifikace testovacích dat v rámci sociálního modelu jsou prezentovány tabulkou 6.7. V počáteční fázi bylo načteno 20 parametrů a zjištěna přesnost klasifikace při optimálním nastavení. Tyto parametry byly navrženy expertem a přesnost klasifikace těchto dat vyjadřuje základní hodnotu následnou analýzu. Postup návrhů vyřazování je stejný jako v předchozích případech. Grafické znázornění posledních výsledků přesností klasifikace vykresluje graf 6.6.

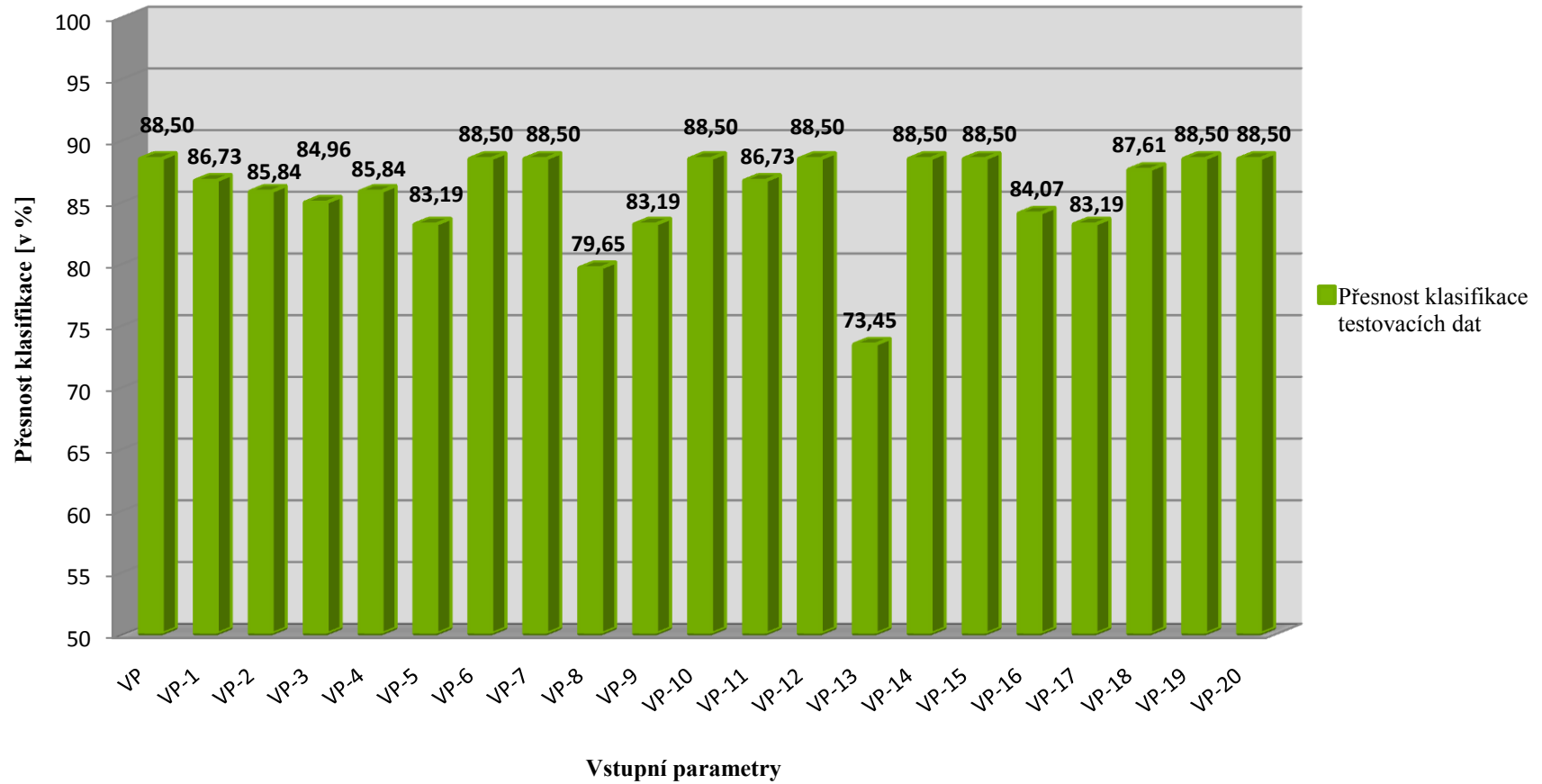
Odchylky přesnosti se pohybují v záporných a nulových hodnotách v intervalu $\langle -15,0443; 0,0000 \rangle$. Celkem osm vyloučených parametrů z celkových dvaceti dosáhlo nulové hodnoty odchylky přesnosti kvalifikace, což znamená, že vyřazením kteréhokoliv z těchto osmi parametrů by nedošlo ke zhoršení výsledků přesností klasifikace.

Pro vybrání parametru pro vyřazení byl brán ohled na vzájemné korelace jednotlivých parametrů, které jsou znázorněny v tabulce 6.14 v příloze 2. Parametr p_{32} , který odpovídá v pořadí desátému parametru, má vysokou závislost se dvěma dalšími parametry. Z tohoto důvodu byl navržen na vyřazení.

Tabulka 6.7: Přesnosti klasifikace sociálních testovacích dat. Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchylna [v %]
VP	1	0	0.0055	true	88,4956	ZH
VP-1	1	0	0.0055	true	86,7257	-1,7699
VP-2	1	0	0.0055	true	85,8407	-2,6549
VP-3	1	0	0.0055	true	84,9558	-3,5398
VP-4	1	0	0.0055	true	85,8407	-2,6549
VP-5	1	0	0.0055	true	83,1858	-5,3098
VP-6	1	0	0.0055	true	88,4956	0,0000
VP-7	1	0	0.0055	true	88,4956	0,0000
VP-8	1	0	0.0055	true	79,6460	-8,8496
VP-9	1	0	0.0055	true	83,1858	-5,3098
VP-10	1	0	0.0055	true	88,4956	0,0000
VP-11	1	0	0.0055	true	86,7257	-1,7699
VP-12	1	0	0.0055	true	88,4956	0,0000
VP-13	1	0	0.0055	true	73,4513	-15,0443
VP-14	1	0	0.0055	true	88,4956	0,0000
VP-15	1	0	0.0055	true	88,4956	0,0000
VP-16	1	0	0.0055	true	84,0708	-4,4248
VP-17	1	0	0.0055	true	83,1858	-5,3098
VP-18	1	0	0.0055	true	87,6106	-0,8850
VP-19	1	0	0.0055	true	88,4956	0,0000
VP-20	1	0	0.0055	true	88,4956	0,0000

Přesnost klasifikace sociálních testovacích dat



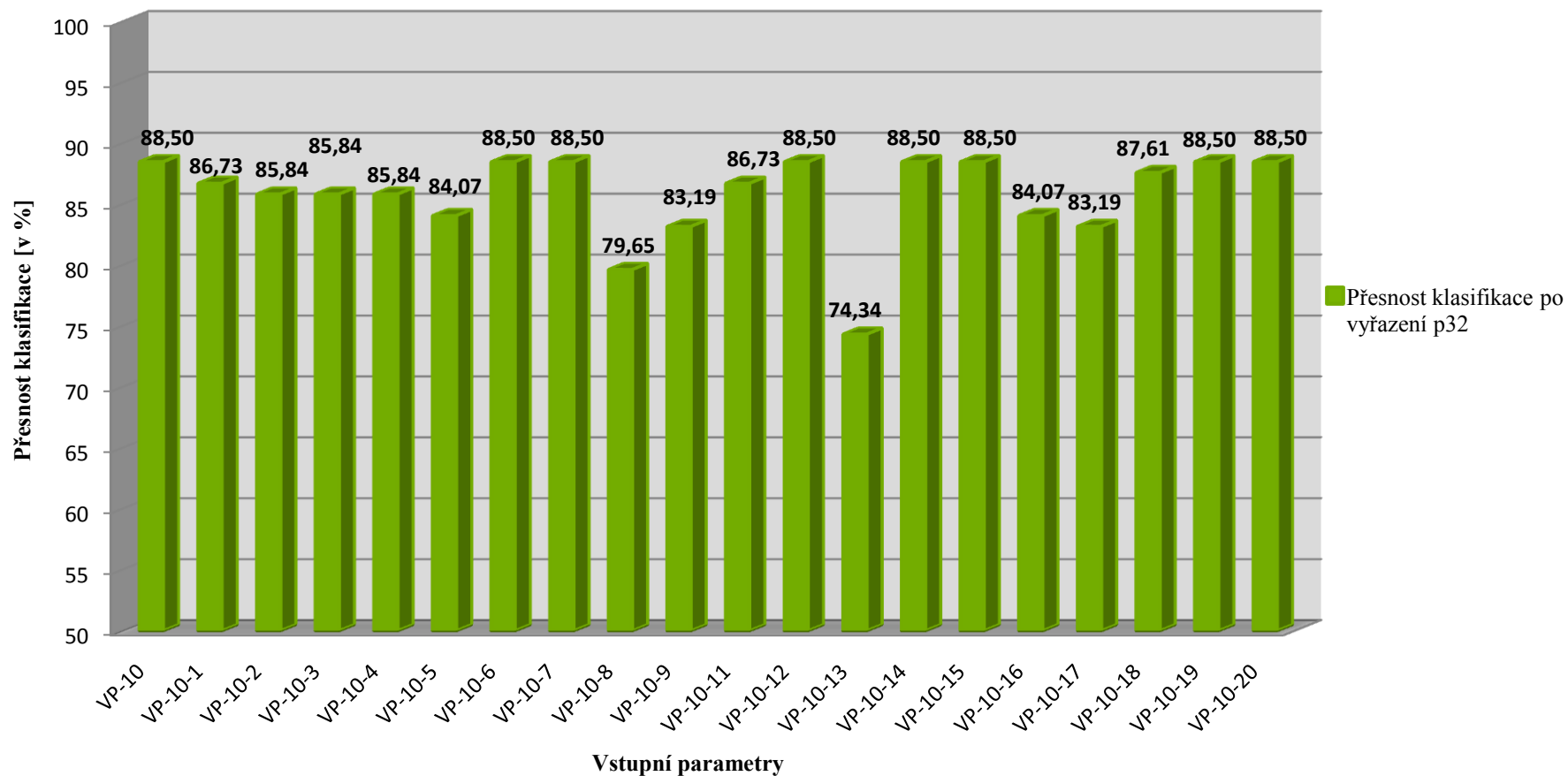
Graf 6.6: Vizualizace přesností klasifikace sociálních testovacích dat. Zdroj: [vlastní]

Po návrhu vyřazení parametru p_{32} byl následující postup prováděn ve stejných krocích. Základní hodnota byla tentokrát vyjádřena jako VK – 10, což představovalo přesnost klasifikace zbylých deseti parametrů po vyloučení v pořadí desátého parametru. Následně bylo zkoumáno, jak ovlivní přesnost klasifikace, bude-li vyřazen další parametr. Opět byly zjištěny přesnosti klasifikací a odchylky přesností. Výsledky prezentuje tabulka 6.8.

Tabulka 6.8: Přesnosti klasifikace po vyloučení parametru p_{32} . Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchylka [v %]
VP-10	1	0	0.0055	true	88,4956	ZH
VP-10-1	1	0	0.0055	true	86,7257	-1,7699
VP-10-2	1	0	0.0055	true	85,8407	-2,6549
VP-10-3	1	0	0.0055	true	85,8407	-2,6549
VP-10-4	1	0	0.0055	true	85,8407	-2,6549
VP-10-5	1	0	0.0055	true	84,0708	-4,4248
VP-10-6	1	0	0.0055	true	88,4956	0,0000
VP-10-7	1	0	0.0055	true	88,4956	0,0000
VP-10-8	1	0	0.0055	true	79,6460	-8,8496
VP-10-9	1	0	0.0055	true	83,1858	-5,3098
VP-10-11	1	0	0.0055	true	86,7257	-1,7699
VP-10-12	1	0	0.0055	true	88,4956	0,0000
VP-10-13	1	0	0.0055	true	74,3363	-14,1593
VP-10-14	1	0	0.0055	true	88,4956	0,0000
VP-10-15	1	0	0.0055	true	88,4956	0,0000
VP-10-16	1	0	0.0055	true	84,0708	-4,4248
VP-10-17	1	0	0.0055	true	83,1858	-5,3098
VP-10-18	1	0	0.0055	true	87,6106	-0,8850
VP-10-19	1	0	0.0055	true	88,4956	0,0000
VP-10-20	1	0	0.0055	true	88,4956	0,0000

Přesnost klasifikace po vyloučení p₃₂



Graf 6.7: Přesnost klasifikace po vyloučení p₃₂. Zdroj: [vlastní]

Odchytky přesnosti se pohybují v záporných a nulových hodnotách v intervalu $(-14,1593; 0,0000)$. Při porovnání odchylek bylo zjištěno, že parametry, které mají nulové hodnoty, odpovídají parametrům, které bylo možné vybrat pro vyloučení už v první části analýzy. Na základě tohoto faktu byl další parametr pro vyřazení vybírán opět podle vzájemných závislostí, které jsou zobrazeny v korelaci sociálních dat v tabulce 6.14 v příloze 2. Pro vyloučení byl navržen v pořadí šestý parametr, který odpovídá p_{28} .

Na grafu 6.7 je znázorněna grafická podoba výsledků přesností kvalifikací pro sociální testovací data po vyřazení parametru p_{32} . Je zde možné vidět, že při vyloučení sedmi různých parametrů dochází k dosažení stejné klasifikace.

Tabulka 6.9: Přesnosti klasifikace po vyloučení parametru p_{32} . Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchytky [v %]
VP-10	1	0	0.0055	true	88,4956	ZH
VP-10-1	1	0	0.0055	true	86,7257	-1,7699
VP-10-2	1	0	0.0055	true	85,8407	-2,6549
VP-10-3	1	0	0.0055	true	85,8407	-2,6549
VP-10-4	1	0	0.0055	true	85,8407	-2,6549
VP-10-5	1	0	0.0055	true	84,0708	-4,4248
VP-10-6	1	0	0.0055	true	88,4956	0,0000
VP-10-7	1	0	0.0055	true	88,4956	0,0000
VP-10-8	1	0	0.0055	true	79,6460	-8,8496
VP-10-9	1	0	0.0055	true	83,1858	-5,3098
VP-10-11	1	0	0.0055	true	86,7257	-1,7699
VP-10-12	1	0	0.0055	true	88,4956	0,0000
VP-10-13	1	0	0.0055	true	74,3363	-14,1593
VP-10-14	1	0	0.0055	true	88,4956	0,0000
VP-10-15	1	0	0.0055	true	88,4956	0,0000
VP-10-16	1	0	0.0055	true	84,0708	-4,4248
VP-10-17	1	0	0.0055	true	83,1858	-5,3098
VP-10-18	1	0	0.0055	true	87,6106	-0,8850
VP-10-19	1	0	0.0055	true	88,4956	0,0000
VP-10-20	1	0	0.0055	true	88,4956	0,0000

Na základě rozsáhlosti a složitosti analýzy, která je způsobena velkým počtem parametrů, a vypovídací schopnosti dat, byla analýza dále zaměřena parametry, které byly v předchozích dvou částech určeny jako možnosti k vyřazení.

V této fázi jsou vyřazeny parametry p_{32} a p_{28} a základní hodnota je VP – 10 – 6. Dále budeme sledovat přesnosti klasifikace v případě vyřazení parametrů na sedmém, dvanáctém, čtrnáctém, patnáctém, devatenáctém a dvacátém pořadí. V případě parametru na dvanáctém a čtrnáctém pořadí došlo vyřazením ke snížení přesnosti, proto dále již nejsou analyzovány. V závislosti na korelacích parametrů byl navržen pro vyřazení devatenáctý parametr, který vyjadřuje parametr p_{41} , viz tabulka 6.9.

Tabulka 6.10: Přesnosti klasifikace po vyloučení parametru p_{28} a p_{32} . Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchylka [v %]
VP-10-6	1	0	0.0055	true	88,4956	ZH
VP-10-6-7	1	0	0.0055	true	88,4956	0,0000
VP-10-6-12	1	0	0.0055	true	87,6106	-0,8850
VP-10-6-14	1	0	0.0055	true	87,6106	-0,8850
VP-10-6-15	1	0	0.0055	true	88,4956	0,0000
VP-10-6-19	1	0	0.0055	true	88,4956	0,0000
VP-10-6-20	1	0	0.0055	true	88,4956	0,0000

Po navržení na vyloučení parametrů p_{32} , p_{28} a p_{41} byla posuzována přesnost klasifikace v případě vyřazení parametrů v pořadí na sedmém, patnáctém a dvacátém místě. V závislosti na korelacích parametrů byl navržen pro vyřazení p_{42} , viz tabulka 6.10.

Tabulka 6.11: Přesnosti klasifikace po vyloučení parametru p_{28} , p_{32} a p_{41} . Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchylka [v %]
VP-10-6-19-7	1	0	0.0055	true	88,4956	0,0000
VP-10-6-19-15	1	0	0.0055	true	88,4956	0,0000
VP-10-6-19-20	1	0	0.0055	true	88,4956	0,0000

V tabulce 6.11 jsou prezentovány výsledky, kdy došlo k navržení vyřazení parametrů p_{32} , p_{28} , p_{41} a p_{42} a sledována přesnost klasifikace při návrhu vyřazení v pořadí sedmého a patnáctého parametru. Na vyřazení byl navržen patnáctý parametr odpovídající p_{37} a poté dle výsledků také sedmý parametr odpovídající p_{29} .

Tabulka 6.12: Přesnosti klasifikace po vyloučení parametru p_{28} , p_{32} , p_{41} a p_{42} . Zdroj: [vlastní]

Vstupní parametry	C	N	L	useRBF	Přesnost klasifikace testovacích dat [v %]	Odchylna [v %]
VP-10-6-19-20-7	1	0	0.0055	true	88,4956	0,0000
VP-10-6-19-20-15	1	0	0.0055	true	88,4956	0,0000
VP-10-6-19-20-15-7	1	0	0.0055	true	88,4956	0,0000

V této části analýzy lze říci, že bylo navrženo na vyřazení pět parametrů a vzhledem k analýze, která byla doposud provedena, je velkou pravděpodobností, že v sociální oblasti by došlo k navržení na odstranění velkého množství parametrů a analýza by byla velmi rozsáhlá. Na základě zjištěných poznatků, proto bylo navrženo vyřazení pouze parametrů p_{28} , p_{29} , p_{32} , p_{41} a p_{42} .

6.6 Zhodnocení výsledků navržených modelů

V rámci ekonomického modelu bylo dosaženo nejlepší přesnosti klasifikace testovacích dat v případě, kdy využití všech původně nastavených parametrů. Přesnost této klasifikace byla 91,1504 %. V ostatních případech byl odebrán vždy pouze jeden parametr a klasifikace těchto dat vycházela vždy nižší než v původním nastavení. Na základě těchto skutečností lze říci, že parametry pro aproximaci byly expertem vhodně navrženy a není navržena žádná změna nastavení parametrů.

Pro environmentální data byl podán návrh na vyřazení tří parametrů. Na základě postupné analýzy bylo dosaženo nejlepší přesnosti klasifikace testovacích dat. Tato přesnost dosáhla hodnoty 94,6903 %. Na základě výsledků byl podán návrh na vyřazení parametrů p_{15} , p_{19} a p_{20} , které tuto přesnost ovlivnily.

V neposlední řadě byla provedena analýza sociálního modelu. V rámci analýzy byly navrženy na vyřazení parametry p_{28} , p_{29} , p_{32} , p_{41} a p_{42} . V případě vyřazení těchto parametrů docházelo k získání stále stejné přesnosti 88,4956 %, což znamená, že jejich vyloučení nemělo žádný vliv na změnu přesnosti.

6.7 Dílčí závěr

Tato kapitola byla věnována aproximaci navržených modelů klasifikátoru parametrů udržitelného rozvoje. Cílem aproximace bylo zjistit, zda parametry pro klasifikaci byly vhodně navrženy. Základem bylo původně nastavených 42 parametrů, které byly vybrány z oblasti ekonomické, environmentální a sociální. Na základě jednotlivých analýz bylo navrženo na vyloučení celkem osm parametrů, z nichž tři byly z oblasti environmentální a pět z oblasti sociální. V rámci ekonomických parametrů nebyl navržen žádný parametr k vyloučení a lze říci, že ekonomické parametry byly vhodně navrženy pro klasifikaci.

ZÁVĚR

Problematika návrhu klasifikátoru parametrů udržitelného rozvoje v lokalitě Pardubického kraje byla řešena pomocí metody Support Vector Machine v programovém prostředí Weka.

V první části diplomové práce byla zpracována problematika udržitelného rozvoje. Byla zavedena definice udržitelného rozvoje a principy a kritéria, které je nutno splňovat. Pro bližší poznání daného tématu byly popsány dvě strategie udržitelného rozvoje, a to strategie navržená Evropskou unií a strategie navržená Českou republikou. V závěru první kapitoly byly definovány ekonomický, environmentální a sociální pilíř a indikátory, jež tvoří základní pohled na udržitelný rozvoj. V rámci teoretického seznámení s metodou Support Vector Machine byl nastíněn úvod do dané problematiky a přiblížena oblast nastavení optimální nadrovinu lineárních separovaných a neseperovaných dat.

V druhé části práce byly navrženy modely pro ekonomické, environmentální a sociální data v programovém prostředí Weka. Optimální hodnoty klasifikátoru byly nastaveny pomocí algoritmu SMO. Na základě analýz byly získány přesnosti klasifikací, které u ekonomických a environmentálních parametrů přesahovaly 90 %, u sociálních dat byla přesnost klasifikace výrazněji nižší.

V závěrečné části diplomové práce v rámci aproximace dat bylo zjišťováno, zda parametry navržené expertem byly vhodně nastaveny. Z výsledků analýzy bylo zjištěno, že ekonomické parametry byly pro klasifikaci navrženy správně a dosahují nejlepších výsledků klasifikace. Celkem bylo na vyřazení navrženo celkem osm parametrů, z nichž tři byly z oblasti environmentální a pět z oblasti sociální.

SEZNAM POUŽITÉ LITERATURY

- [1] BUKOVIČ, L. *Modelování bonity obcí pomocí Support Vector Machines*. Pardubice, 2008. 67 s. Diplomová práce. Univerzita Pardubice, Fakulta ekonomicko-správní.
- [2] CENIA, česká informační agentura životního prostředí [online]. 2009 [cit. 2010-02-20]. Udržitelný rozvoj. Dostupné z WWW: <[http://www.cenia.cz/web/www/web-pub2.nsf/\\$pid/MZPMSFHV0HSB](http://www.cenia.cz/web/www/web-pub2.nsf/$pid/MZPMSFHV0HSB)>.
- [3] Computer Science Department, University of Waikato [online]. 2009 [cit. 2009-11-15]. Weka 3 - Data Mining with Open Source Machine Learning Software in Java. Dostupné z WWW: <<http://www.cs.waikato.ac.nz/ml/weka/>>.
- [4] Český statistický úřad, Krajská správa Pardubice. *Vybrané oblasti udržitelného rozvoje v Pardubickém kraji*. Pardubice : Český statistický úřad, 2007. 154 s. Dostupné z WWW: <<http://czso.cz/xed/edicniplan.nsf/p/13-53n39-07>>. ISBN 978-80-250-1621-3.
- [5] Český statistický úřad [online]. 2010 [cit. 2010-04-06]. Sledované jevy, použité ukazatele a jejich metodika. Dostupné z WWW: <[http://www.czso.cz/csu/redakce.nsf/i/sledovane_jevy_pouzite_ukazatele_a_jejich_metodika/\\$File/jevy_ukaz_uap_100324.htm](http://www.czso.cz/csu/redakce.nsf/i/sledovane_jevy_pouzite_ukazatele_a_jejich_metodika/$File/jevy_ukaz_uap_100324.htm)>.
- [6] HÁJEK, P. *Modelování bonity obcí metodami výpočetní inteligence*. Pardubice, 2006, 171 s. Disertační práce. Univerzita Pardubice, Fakulta ekonomicko-správní.
- [7] HÁJEK, P., OLEJ, V. *Municipal Creditworthiness Modelling by Kernel-based Approaches with Supervised and Semi-supervised Learning*. Proc. of the 11th International Conference on Engineering Applications of Neural Networks, EANN 2009, Communications in Computer and Information Science, Engineering Applications of Neural Networks, London, Palmer-Brown, D., Draganova, Ch., Pimenidis, E., Mouratidis, H., Eds., 27-29 august, Springer Berlin Heidelberg New York, 2009, pp.35-44, ISSN 1865-0929, ISBN 978-3-642-03968-3.

- [8] HAYKIN, S. *Neural networks : a comprehensive foundation*. Upper Saddle River : Prentice Hall, 1999, 842 s. ISBN 0-13-273350-1.
- [9] HŘEBÍK, Š., TŘEBICKÝ, V. *Budoucí rozvoj Ústeckého kraje (BROÚK)* [online]. Praha : EnviConsult, s.r.o., 2006 [cit. 2010-04-19]. 48 s. Manuál zpracování a využití sady indikátorů rozvoje pro malé obce. Dostupné z WWW: <http://brouk.kr-ustecky.cz/soubory/450018/manual_str.01-48_b5.pdf>.
- [10] *Informační systém Masarykovy univerzity* [online]. 2005 [cit. 2010-02-23]. Support Vector Machine. Dostupné z WWW: <http://is.muni.cz/el/1433/podzim2006/PA034/09_SVM.pdf?fakulta=1433;obdobi=3523;kod=PA034>.
- [11] JENÍČEK, V., FOLTÝN, J. *Globální problémy a světová ekonomika*. Praha : C.H. Beck, 2003. 269 s. ISBN 80-7179-795-2.
- [12] KELBEL, J., ŠILHÁN, D. *Gerster laboratory* [online]. 2005 [cit. 2010-03-15]. Shluková analýza. Dostupné z WWW: <<http://gerstner.felk.cvut.cz/biolab/X33BMI/slides/KMeans.pdf>>.
- [13] KUBANOVÁ, J. *Statistické metody pro ekonomickou a technickou praxi*. Bratislava : Statis, 2003. 247 s. ISBN 80-85659-31-X.
- [14] KVASNIČKA, V. et al. *Úvod do teórie neurónových sietí*. Bratislava : IRIS, 1997. 258 s.
- [15] MACH, M. *Mikroekonomie II*. Vyd. 3. Slaný : Melandrium, 2001. 367 s. ISBN 80-86175-18-9.
- [16] *Ministerstvo práce a sociálních věcí ČR* [online]. 2005 [cit. 2010-04-19]. 137 s. Formulace doporučení MPSV v oblasti sociálního pilíře udržitelného rozvoje vycházejících z mezinárodní komparace. Dostupné z WWW: <http://www.mpsv.cz/files/clanky/2902/Zaverecna_zprava_2005.pdf>.
- [17] *Ministerstvo práce a sociálních věcí ČR* [online]. 2005 [cit. 2010-04-01]. Zahraniční rozvojová spolupráce České republiky. Dostupné z WWW: <<http://www.mpsv.cz/cs/1605>>.

- [18] MOLDAN, B. *Ekologická dimenze udržitelného rozvoje*. Praha : Karolinum, 2001. 102 s. ISBN 80-246-0246-6.
- [19] *Pardubický kraj* [online]. 2008 [cit. 2010-03-10]. O kraji. Dostupné z WWW: <<http://www.pardubickykraj.cz/index.asp?thema=2613&category>>.
- [20] PLATT, J. *Microsoft research* [online]. 1998 [cit. 2010-03-15]. Shluková analýza. Dostupné z WWW: <<http://research.microsoft.com/en-us/um/people/jplatt/smoTR.pdf>>.
- [21] ŠILHÁNKOVÁ, V. *Indikátory udržitelného rozvoje : teoretické přístupy a zkušenosti v České republice : Případová studie Hradec Králové*. Hradec Králové : Civitas per Populi, 2008. 52 s. ISBN 978-80-903813-6-0.
- [22] ŠUŠLÍKOVÁ, M. *Unesco Kroměříž* [online]. 2009 [cit. 2010-04-05]. Výchova a osvěta – klíč k udržitelnému rozvoji. Dostupné z WWW: <http://www.unesco-kromeriz.cz/udrzitelny_rozvoj/co_je_udrzitelny_rozvoj.pdf>.
- [23] *Týmová iniciativa pro místní udržitelný rozvoj* [online]. 2005 [cit. 2010-03-11]. Co je indikátor. Dostupné z WWW: <<http://www.timur.cz/indikatory/co-je-indikator.html>>.
- [24] *Univerzita Karlova v Praze* [online]. 2005 [cit. 2010-04-01]. Co jsou to indikátory. Dostupné z WWW: <<http://www.cuni.cz/COZP-14.html>>.
- [25] *Ústav územního rozvoje* [online]. 2009 [cit. 2010-03-20]. Pilíře udržitelného rozvoje. Dostupné z WWW: <http://www.uur.cz/images/pap/KapitolaA/A12_PilireUdrzitelnehoRozvoje_20090710.pdf>.
- [26] *Wikipedia* [online]. 2008 [cit. 2010-03-19]. Support vector machine. Dostupné z WWW: <http://en.wikipedia.org/wiki/Support_vector_machine>.
- [27] Zach, O. *Modelování bonity obcí pomocí Support Vector Machines*. Pardubice, 2008. 71 s. Diplomová práce. Univerzita Pardubice, Fakulta ekonomicko-správní.

SEZNAM ZKRATEK

C	The complexity constant
ČR	Česká republika
eko	ekonomický
envi	environmentální
EU	Evropská unie
KP	Kvadratické programování
L	The tolerance parameter
N	filter Type
OSN	Organizace spojených národů
RBF	Radial Basic Function
SMO	Sequential Minimal Optimization
soc	sociální
SVM	Support Vector Machine
VP	Všechny faktory
WCED	The World Commission on Environment and Development (Světová komise při životní prostředí a rozvoj)
ZH	Základní hodnota
ZRS	Zahraniční rozvojová spolupráce

SEZNAM OBRÁZKŮ

Obrázek 2.1: Architektura Support Vector Machine. Zdroj: [7].....	19
Obrázek 2.2: Optimální nadrovina pro lineárně separovaná data. Zdroj: [8].....	20
Obrázek 2.3: Optimální nadrovina pro neseparovaná data I. Zdroj: [8]	22
Obrázek 2.4: Optimální nadrovina pro neseparovaná data II. Zdroj: [8].....	23
Obrázek 3.1: Pardubický kraj. Zdroj: [19]	24
Obrázek 4.1: Návrh modelu pro klasifikaci obcí. Zdroj: [vlastní]	36
Obrázek 4.2: Rozdělení objektů do stanovených tříd. Zdroj: [vlastní]	38
Obrázek 4.3: Rozklad objektů na množinu testovací a trénovací. Zdroj: [14].....	40
Obrázek 4.4: Návrh modelu v prostředí Weka. Zdroj: [vlastní]	42
Obrázek 5.1: Nastavení záhlaví pro import dat. Zdroj: [vlastní].....	44
Obrázek 5.2: Výsledek klasifikace ekonomických dat. Zdroj: [vlastní]	47
Obrázek 5.3: Výsledek klasifikace environmentálních dat. Zdroj: [vlastní].....	50
Obrázek 5.4: Výsledek klasifikace sociálních dat. Zdroj: [vlastní]	54

SEZNAM GRAFŮ

Graf 5.1: Očekávané a skutečné rozdělení ekonomických dat. Zdroj: [vlastní]	48
Graf 5.2: Očekávané a skutečné rozdělení environmentálních dat. Zdroj: [vlastní]	51
Graf 5.3: Očekávané a skutečné rozdělení sociálních dat. Zdroj: [vlastní].....	55
Graf 6.1: Vizualizace přesností klasifikace ekonomických testovacích dat. Zdroj: [vlastní]	61
Graf 6.2: Vizualizace přesností klasifikace environmentálních testovacích dat. Zdroj: [vlastní]	62
Graf 6.3: Přesnost klasifikace po vyloučení p_{15} . Zdroj: [vlastní]	64
Graf 6.4: Přesnost klasifikace po vyloučení p_{15} a p_{20} . Zdroj: [vlastní].....	65
Graf 6.5: Přesnost klasifikace po vyloučení p_{15} , p_{19} a p_{20} . Zdroj: [vlastní]	66
Graf 6.6: Vizualizace přesností klasifikace sociálních testovacích dat. Zdroj: [vlastní].....	69
Graf 6.7: Přesnost klasifikace po vyloučení p_{32} . Zdroj: [vlastní]	71

SEZNAM TABULEK

Tabulka 3.1: Souhrn ekonomický indikátorů.....	27
Tabulka 3.2: Souhrn environmentálních indikátorů.....	29
Tabulka 3.3: Souhrn sociálních indikátorů	34
Tabulka 4.1: Popisná statistika ekonomický indikátorů. Zdroj: [vlastní]	85
Tabulka 4.2: Popisná statistika environmentálních indikátorů. Zdroj: [vlastní]	85
Tabulka 4.3: Popisná statistika sociálních indikátorů. Zdroj: [vlastní].....	86
Tabulka 5.1: Důležité parametry algoritmu SMO Zdroj: [vlastní]	45
Tabulka 5.2: Návrh parametrů modelu pro ekonomická data I. Zdroj: [vlastní]	45
Tabulka 5.3: Návrh parametrů modelu pro ekonomická data II. Zdroj: [vlastní].....	46
Tabulka 5.4: Optimální struktura modelu klasifikace pro ekonomická data. Zdroj: [vlastní]	46
Tabulka 5.5: Porovnání očekávaného a skutečného rozložení eko dat. Zdroj: [vlastní].....	48
Tabulka 5.6: Návrh parametrů modelu pro environmentální data I. Zdroj: [vlastní].....	49
Tabulka 5.7: Návrh parametrů modelu pro environmentální data II. Zdroj: [vlastní]	49
Tabulka 5.8: Optimální struktura modelu klasifikace pro envi data. Zdroj: [vlastní].....	50
Tabulka 5.9: Porovnání očekávaného a skutečného rozložení envi dat. Zdroj: [vlastní].....	52
Tabulka 5.10: Návrh parametrů modelu pro sociální data I. Zdroj: [vlastní]	53
Tabulka 5.11: Návrh parametrů modelu pro sociální data II. Zdroj: [vlastní]	53
Tabulka 5.12: Optimální struktura modelu klasifikace pro sociální data. Zdroj: [vlastní]	53
Tabulka 5.13: Porovnání očekávaného a skutečného rozložení soc dat. Zdroj: [vlastní]	55
Tabulka 6.1: Ukázkový příklad prezentace výsledků analýzy. Zdroj: [vlastní].....	58
Tabulka 6.2: Přesnosti klasifikace ekonomických testovacích dat. Zdroj: [vlastní]	60
Tabulka 6.3: Přesnosti klasifikace environmentálních testovacích dat. Zdroj: [vlastní].....	62
Tabulka 6.4: Přesnosti klasifikace po vyloučení parametru p_{15} . Zdroj: [vlastní].....	63
Tabulka 6.5: Přesnosti klasifikace po vyloučení parametru p_{15} a p_{20} . Zdroj: [vlastní].....	65
Tabulka 6.6: Přesnosti klasifikace po vyloučení parametru p_{15} , p_{19} a p_{20} . Zdroj: [vlastní]	66
Tabulka 6.7: Přesnosti klasifikace sociálních testovacích dat. Zdroj: [vlastní]	68
Tabulka 6.8: Přesnosti klasifikace po vyloučení parametru p_{32} . Zdroj: [vlastní].....	70
Tabulka 6.9: Přesnosti klasifikace po vyloučení parametru p_{32} . Zdroj: [vlastní]	72
Tabulka 6.10: Přesnosti klasifikace po vyloučení parametru p_{28} a p_{32} . Zdroj: [vlastní].....	73
Tabulka 6.11: Přesnosti klasifikace po vyloučení parametru p_{28} , p_{32} a p_{41} . Zdroj: [vlastní]	73
Tabulka 6.12: Přesnosti klasifikace po vyloučení parametru p_{28} , p_{32} , p_{41} a p_{42} . Zdroj: [vlastní].	74
Tabulka 6.13: Korelace ekonomických indikátorů. Zdroj: [vlastní]	87
Tabulka 6.14: Korelace environmentálních indikátorů. Zdroj: [vlastní].....	87
Tabulka 6.15: Korelace sociálních indikátorů. Zdroj: [vlastní]	88

SEZNAM PŘÍLOH

příloha 1 - Popisné statistiky dat

příloha 2 – Korelace indikátorů

příloha 1 - Popisné statistiky dat

Tabulka 4.1: Popisná statistika ekonomický indikátorů. Zdroj: [vlastní]

<i>Parametr</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Průměr</i>	<i>Medián</i>	<i>Směrodatná odchylka</i>
P1	32,14	86,36	58,37	58,97	9,61
P2	0,11	0,44	0,20	0,19	0,05
P3	0,93	1,09	0,99	0,95	0,06
P4	0,00	0,82	0,07	0,03	0,11
P5	0,00	306 480,82	4 179,54	1 262,03	15 809,77
P6	0,26	2,59	1,21	1,18	0,36
P7	0,00	0,62	0,08	0,06	0,08
P8	0,00	0,83	0,22	0,19	0,18
P9	5 422,78	820 591,81	69 577,01	55 281,35	65 270,54
P10	74,89	63 499,69	7 064,03	3 429,48	10 787,33
P11	43,81	761,63	199,10	193,99	57,46

Tabulka 4.2: Popisná statistika environmentálních indikátorů. Zdroj: [vlastní]

<i>Parametr</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Průměr</i>	<i>Medián</i>	<i>Směrodatná odchylka</i>
P12	0,00	1,00	0,26	0,00	0,44
P13	0,00	1,00	0,39	0,00	0,49
P14	2,10	97,60	73,17	77,50	17,29
P15	2,50	55,10	8,37	6,90	5,79
P16	0,00	14,60	1,42	0,70	1,94
P17	0,03	13,78	1,19	0,74	1,57
P18	0,00	0,16	0,05	0,04	0,02
P19	0,00	0,27	0,01	0,00	0,02
P20	0,00	2,21	0,29	0,21	0,30
P21	0,00	8,78	0,74	0,39	1,09
P22	0,05	6,15	1,33	1,18	0,80

Tabulka 4.3: Popisná statistika sociálních indikátorů. Zdroj: [vlastní]

<i>Parametr</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Průměr</i>	<i>Medián</i>	<i>Směrodatná odchylka</i>
P23	0,06	0,46	0,26	0,27	0,07
P24	0,01	0,22	0,11	0,12	0,05
P25	0,00	1,08	0,11	0,07	0,12
P26	0,00	0,40	0,02	0,00	0,04
P27	0,16	0,67	0,35	0,34	0,05
P28	0,00	0,72	0,07	0,05	0,08
P29	0,00	100,00	96,94	98,81	11,09
P30	0,00	96,60	33,98	34,87	32,17
P31	0,00	99,54	22,64	10,43	28,43
P32	27,00	88 181,00	1 120,72	340,00	4 672,83
P33	0,00	50,00	11,71	10,47	5,88
P34	0,00	16,18	3,10	2,86	1,93
P35	0,00	1,00	0,25	0,00	0,43
P36	-53,00	131,00	1,27	0,00	9,42
P37	-129,00	555,00	6,82	2,00	38,63
P38	3,30	30,10	15,27	15,40	3,09
P39	5,99	36,36	15,13	14,28	4,23
P40	0,00	0,14	0,04	0,03	0,02
P41	0,00	0,10	0,00	0,00	0,01
P42	0,00	1,00	0,00	0,00	0,05

příloha 2 – Korelace indikátorů

Tabulka 6.13: Korelace ekonomických indikátorů. Zdroj: [vlastní]

	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀	P ₁₁
P ₁	1,00	0,74	-0,26	0,03	-0,03	0,13	0,15	-0,01	-0,05	-0,25	-0,20
P ₂	0,74	1,00	-0,20	0,06	-0,04	-0,01	0,03	0,01	-0,09	-0,07	-0,26
P ₃	-0,26	-0,20	1,00	0,02	0,06	0,09	-0,07	-0,04	0,00	-0,04	0,30
P ₄	0,03	0,06	0,02	1,00	0,15	0,02	-0,04	0,11	0,06	0,03	-0,06
P ₅	-0,03	-0,04	0,06	0,15	1,00	0,00	0,05	0,06	0,38	0,08	0,04
P ₆	0,13	-0,01	0,09	0,02	0,00	1,00	0,09	0,42	-0,12	-0,56	-0,03
P ₇	0,15	0,03	-0,07	-0,04	0,05	0,09	1,00	0,00	0,34	-0,18	-0,04
P ₈	-0,01	0,01	-0,04	0,11	0,06	0,42	0,00	1,00	0,09	0,04	-0,03
P ₉	-0,05	-0,09	0,00	0,06	0,38	-0,12	0,34	0,09	1,00	0,22	0,14
P ₁₀	-0,25	-0,07	-0,04	0,03	0,08	-0,56	-0,18	0,04	0,22	1,00	0,05
P ₁₁	-0,20	-0,26	0,30	-0,06	0,04	-0,03	-0,04	-0,03	0,14	0,05	1,00

Tabulka 6.14: Korelace environmentálních indikátorů. Zdroj: [vlastní]

	P ₁₂	P ₁₃	P ₁₄	P ₁₅	P ₁₆	P ₁₇	P ₁₈	P ₁₉	P ₂₀	P ₂₁	P ₂₂
P ₁₂	1,00	-0,47	-0,05	0,35	0,17	0,02	-0,34	-0,03	-0,14	-0,12	-0,35
P ₁₃	-0,47	1,00	0,08	-0,13	-0,06	-0,09	0,03	0,03	-0,02	-0,05	0,09
P ₁₄	-0,05	0,08	1,00	0,04	0,05	-0,72	-0,10	0,06	-0,63	-0,33	0,11
P ₁₅	0,35	-0,13	0,04	1,00	0,21	-0,20	-0,37	-0,03	-0,25	-0,30	-0,39
P ₁₆	0,17	-0,06	0,05	0,21	1,00	-0,11	-0,21	-0,08	-0,12	-0,18	-0,19
P ₁₇	0,02	-0,09	-0,72	-0,20	-0,11	1,00	0,16	-0,04	0,53	0,71	-0,02
P ₁₈	-0,34	0,03	-0,10	-0,37	-0,21	0,16	1,00	0,25	0,50	0,43	0,72
P ₁₉	-0,03	0,03	0,06	-0,03	-0,08	-0,04	0,25	1,00	-0,01	0,04	0,23
P ₂₀	-0,14	-0,02	-0,63	-0,25	-0,12	0,53	0,50	-0,01	1,00	0,58	0,56
P ₂₁	-0,12	-0,05	-0,33	-0,30	-0,18	0,71	0,43	0,04	0,58	1,00	0,37
P ₂₂	-0,35	0,09	0,11	-0,39	-0,19	-0,02	0,72	0,23	0,56	0,37	1,00

Tabulka 6.15: Korelace sociálních indikátorů. Zdroj: [vlastní]

	P ₂₃	P ₂₄	P ₂₅	P ₂₆	P ₂₇	P ₂₈	P ₂₉	P ₃₀	P ₃₁	P ₃₂	P ₃₃	P ₃₄	P ₃₅	P ₃₆	P ₃₇	P ₃₈	P ₃₉	P ₄₀	P ₄₁	P ₄₂
P ₂₃	1,00	0,50	-0,18	-0,35	0,07	-0,02	-0,10	0,07	-0,23	-0,36	-0,17	-0,16	-0,28	-0,21	-0,11	-0,05	0,02	-0,11	-0,01	0,01
P ₂₄	0,50	1,00	-0,12	-0,48	-0,10	0,02	-0,08	-0,12	-0,29	-0,30	0,06	-0,26	-0,35	-0,17	-0,14	-0,08	-0,02	0,01	-0,02	-0,04
P ₂₅	-0,18	-0,12	1,00	0,25	0,05	-0,06	0,08	0,11	0,19	0,09	-0,12	0,24	0,12	0,10	0,11	-0,04	-0,05	-0,09	-0,01	0,04
P ₂₆	-0,35	-0,48	0,25	1,00	0,01	-0,13	0,01	0,11	0,31	0,18	-0,05	0,26	0,16	0,14	0,07	0,02	-0,11	-0,03	-0,06	0,05
P ₂₇	0,07	-0,10	0,05	0,01	1,00	0,50	-0,08	-0,05	-0,03	0,08	0,05	0,09	-0,01	-0,01	0,00	-0,31	0,63	0,18	-0,10	-0,02
P ₂₈	-0,02	0,02	-0,06	-0,13	0,50	1,00	-0,12	-0,35	-0,31	-0,14	0,27	-0,21	-0,22	-0,13	-0,10	-0,17	0,52	0,27	-0,06	-0,04
P ₂₉	-0,10	-0,08	0,08	0,01	-0,08	-0,12	1,00	0,09	-0,06	0,03	-0,12	0,04	0,09	0,05	0,01	0,09	-0,06	-0,07	0,03	0,01
P ₃₀	0,07	-0,12	0,11	0,11	-0,05	-0,35	0,09	1,00	0,33	0,19	-0,31	0,27	0,10	0,14	0,13	-0,10	-0,09	-0,27	0,16	0,03
P ₃₁	-0,23	-0,29	0,19	0,31	-0,03	-0,31	-0,06	0,33	1,00	0,34	-0,23	0,36	0,26	0,20	0,16	0,00	-0,21	-0,20	0,02	0,05
P ₃₂	-0,36	-0,30	0,09	0,18	0,08	-0,14	0,03	0,19	0,34	1,00	-0,09	0,34	0,26	0,71	0,58	-0,04	-0,05	-0,06	0,02	0,02
P ₃₃	-0,17	0,06	-0,12	-0,05	0,05	0,27	-0,12	-0,31	-0,23	-0,09	1,00	-0,35	-0,06	-0,08	-0,12	0,03	0,13	0,72	-0,11	-0,05
P ₃₄	-0,16	-0,26	0,24	0,26	0,09	-0,21	0,04	0,27	0,36	0,34	-0,35	1,00	0,19	0,25	0,19	-0,06	-0,07	-0,27	0,07	0,15
P ₃₅	-0,28	-0,35	0,12	0,16	-0,01	-0,22	0,09	0,10	0,26	0,26	-0,06	0,19	1,00	0,15	0,12	-0,02	-0,14	-0,05	0,06	-0,03
P ₃₆	-0,21	-0,17	0,10	0,14	-0,01	-0,13	0,05	0,14	0,20	0,71	-0,08	0,25	0,15	1,00	0,53	0,06	-0,13	-0,05	0,07	-0,01
P ₃₇	-0,11	-0,14	0,11	0,07	0,00	-0,10	0,01	0,13	0,16	0,58	-0,12	0,19	0,12	0,53	1,00	0,01	-0,05	-0,08	0,27	0,02
P ₃₈	-0,05	-0,08	-0,04	0,02	-0,31	-0,17	0,09	-0,10	0,00	-0,04	0,03	-0,06	-0,02	0,06	0,01	1,00	-0,35	0,02	0,10	0,03
P ₃₉	0,02	-0,02	-0,05	-0,11	0,63	0,52	-0,06	-0,09	-0,21	-0,05	0,13	-0,07	-0,14	-0,13	-0,05	-0,35	1,00	0,09	-0,10	-0,04
P ₄₀	-0,11	0,01	-0,09	-0,03	0,18	0,27	-0,07	-0,27	-0,20	-0,06	0,72	-0,27	-0,05	-0,05	-0,08	0,02	0,09	1,00	-0,03	-0,04
P ₄₁	-0,01	-0,02	-0,01	-0,06	-0,10	-0,06	0,03	0,16	0,02	0,02	-0,11	0,07	0,06	0,07	0,27	0,10	-0,10	-0,03	1,00	-0,02
P ₄₂	0,01	-0,04	0,04	0,05	-0,02	-0,04	0,01	0,03	0,05	0,02	-0,05	0,15	-0,03	-0,01	0,02	0,03	-0,04	-0,04	-0,02	1,00