

Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Katedra matematiky

Semestrální práce KMA/AGI

**Aplikace vyhledání regionálních vzorů
pomocí prostorového dolování dat
v SŘBD Oracle**

31. 5. 2010

Bc. Hana Hůlová

Úvod

Tato semestrální práce byla zpracována v rámci mé diplomové práce nesoucí název: *Aplikace vybraných metod prostorového dolování dat v databázových systémech*. Úkolem semestrální práce bylo aplikovat metodu asociace, nebo-li vyhledávání regionálních vzorů v podobě asociačních pravidel, na reálná data, které poskytl Preventivně informační odbor Policejního prezidia ČR. V první kapitole je uvedeno několik definic prostorového dolování dat a popis procesu prostorového dolování. Ve druhé kapitole je detailně popsáno vlastní dolování skrytých závislostí a vzorů na konkrétních datech a ve třetí kapitole je provedeno zhodnocení výsledků, tedy vygenerovaných asociačních pravidel pro jednotlivé kraje ČR a plošně přes celou ČR.

1 Prostorové dolování dat (Spatial Data Mining)

Nástup dálkového průzkumu Země a jiných technologií v posledním desetiletí výrazně zvýšil možnosti sběru velkého množství geografických dat. S tím je spojena potřeba nových metod a nástrojů, které mohou automaticky a inteligentně transformovat geografická *data* do *informací* a dále spojovat geografické znalosti. To vyžaduje nové přístupy v geografické reprezentaci, ve zpracování dotazů, v prostorových analýzách a ve vizualizaci dat.

1.1 Definice

V této části jsou vypsány vybrané definice pojmu prostorové dolování dat.

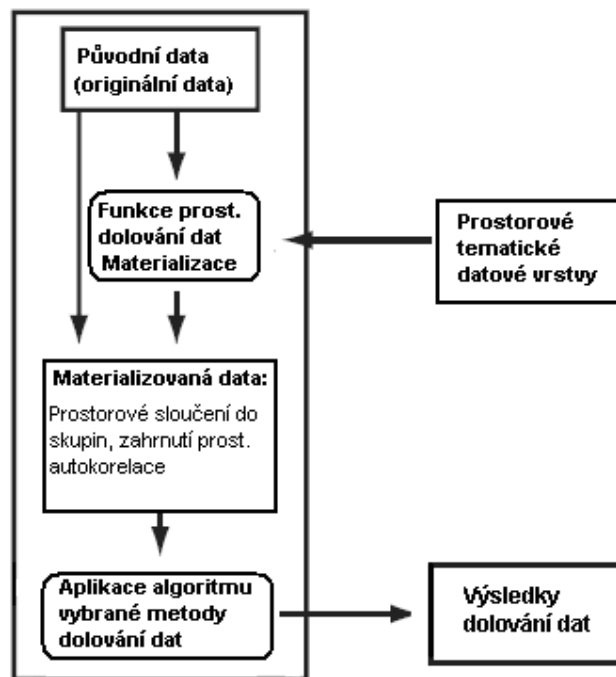
„Spatial Data Mining (SDM) is a well identified domain of data mining. It can be defined as the discovery of interesting, implicit and previously unknown knowledge from large spatial data bases.“ [6]

„Spatial Data Mining extends relational data mining with respect to special features of spatial data, like mutual influence of neighboring objects by certain factors (topology, distance, direction).“ [6]

„Dolování prostorových dat je proces objevování zajímavých, dosud neznámých, ale potenciálně užitečných zákonitostí a poznatků v prostorových databázích.“ [3]

1.2 Proces prostorového dolování dat

Do procesu dolování dat z relačních i z prostorových databází je nutné zahrnout velmi důležitou fázi, kterou je příprava dat. V prostorovém dolování dat se jedná o fázi, ve které je do původních dat zahrnut vliv okolí, tedy prostorové vlastnosti a relační vztahy s okolními prvky. Pro tento krok je používán termín materializace (*materialization*), a získaná, předpřipravená data jsou poté materializovaná data (*materialized data*). Na takto připravená data je poté možné přímo aplikovat některou z metod dolování dat, které jsou v mé diplomové práci popsány v kapitole 3.1. Jednotlivé kroky procesu prostorového dolování dat jsou uvedeny na Obrázku 1.



Obr. 1: Fáze prostorového dolování dat

1.2.1 Původní data

Vstupem do procesu prostorového dolování dat jsou původní, originální data, získaná sběrem dat, obsahující prostorové i neprostorové informace o sledovaném jevu.

V klasickém dolování dat by těmito údaji mohly být např. údaje o telefonních hovorech určité skupiny lidí, kdy, s kým a jak často proběhly. V případě prostorového dolování dat už musí být zahrnuta i prostorová informace, např. souřadnice nějakého sledovaného jevu. Jako příklad mohou být uvedeny dopravní nehody. Při nich policie eviduje jak souřadnice místa nehody, tak evidenční údaje o účastnících, autech aj.

1.2.2 Funkce prostorového dolování dat

Prostorová informace v původních datech je zpracována funkcemi prostorového dolování dat, aby byla získána materializovaná data. Mezi tyto funkce patří [1]:

- *Prostorové sloučení (Spatial Binning)* - rozděluje objekty (např. místa dopravních nehod) podle jejich prostorové informace do skupin, tzv. binů (*bins*), vytvořených podle geografických areálů (např. krajů ČR). Tato funkce je podrobně popsána a aplikována na konkrétní data v kapitole 5.4. Data připravená touto funkcí je možné využít k získávání zajímavých regionálních vzorů nebo hledání asociačních pravidel[1].
- *Prostorová korelace (Materializing Spatial Correlation)* - Umístění konkrétního objektu v prostoru je ovlivněno jak neprostorovými atributy objektu samotného, tak atributy okolních objektů. Princip této funkce spočívá v tom, že je do původní tabulky přidán sloupec, který uvažuje i hodnoty okolních prvků. Pokud zůstaneme u výše zmíněného příkladu dopravních nehod, bude tento sloupec obsahovat např. počet dopravních nehod v okolí 10 km od našeho konkrétního objektu[1].

Takto získaná data lze využít opět k vyhledání asociačních pravidel anebo na data lze aplikovat metodu shlukování, která odhalí místa s největším nebezpečím dopravních nehod.

1.2.3 Aplikace vybrané metody dolování dat

Do této fáze procesu přichází již materializovaná data, což jsou data ve formě, kterou vyžaduje daná jednotlivá metoda dolování. Jako příklad zde může být uvedena metoda asociace, která vyžaduje, aby byla původní data rozdělena do skupin. V následující tabulce jsou uvedeny metody prostorového dolování dat a algoritmy, které se s jednotlivými metodami pojí.

Tab. 1: Metody dolování dat a jejich algoritmy

Metody dolování dat	Algoritmy
regrese	Generalizovaný lineární model
klasifikace	Rozhodovací stromy, Naive Bayes
shlukování	K-means
asociace	Apriori

Výstupem této části jsou již výsledky dolování, tedy zajímavé znalosti a zákonitosti, které nejsou vidět v databázích na první pohled. Tyto výsledky mohou být velmi cenné, např. pokud by šlo o vyhledání míst s největší koncentrací dopravních nehod, mohou tato data posloužit k návržení lepšího značení inkriminovaného úseku silnice, nebo pro zavedení snížené rychlosti, nebo pro umístění radaru aj.

2 Praktické ověření

Pro zpracování reálných dat byly použity nástroje relačního databázového systému Oracle, který velmi dobře podporuje prostorové dolování dat. Z výše uvedených metod prostorového dolování dat byla pro praktické ověření vyhledání závislostí v konkrétních datech vybrána metoda asociace. Vzhledem k charakteru údajů o trestných činech, které byly poskytnuty Preventivně informačním odborem Policejního prezidia ČR, se použití této metody jeví jako účelné. Cílem bylo vyhledat regionální vzory v podobě asociačních pravidel, platné v jednotlivých krajích ČR, a porovnat je se vzory, platnými pro celou ČR.

2.1 Vymezení problému

Prvotním úkolem před vlastním řešením dané úlohy bylo vymezit si problém, porozumět cíli a požadavkům, kterých by mělo být prostorovým dolováním dat dosaženo. Z mnoha údajů o trestných činech byla jako cíl zájmu vybrána kriminalita dětí a mladistvých, přičemž se jednalo o trestný čin loupeže.

Bylo položeno několik otázek, na něž byly hledány odpovědi právě metodou asociace:

- Souvisí kriminalita dětí a mladistvých s krajem, ve kterém žijí?
- Jak se liší asociační pravidla vygenerovaná pro celou Českou republiku od pravidel pro jednotlivé kraje?
- Je nějaká souvislost mezi dnem v týdnu a spácháním loupeže dětmi a mladistvými?
- Spáchají děti loupež spíše ve dne nebo v noci?
- Pokud spáchá trestný čin loupeže mládež, jaký podíl v celkovém součtu mají děti a jaký mladiství?

Kromě výše uvedených otázek mohou být vyhledána i další asociační pravidla, která budu z pohledu řešeného problému zajímavá.

2.2 Sběr dat

Data pro zpracování této práce byla poskytnuta Policií ČR, přičemž se jednalo o tabulky, vytvořené v programu Microsoft Excel, zaznamenávající všechny trestné činy od roku 2000 do roku 2008. V datech bylo možné zjistit počet vražd, znásilnění, ublížení na zdraví, loupeží, krádeží aut a jiných trestných činů, dále bylo uvedeno, zda byly činy objasněny, ve který den bylo kolik trestných činů spácháno a v jakém časovém úseku během dne. Jako poslední informací v tabulkách bylo uvedeno, kdo loupež spáchal, zda recidivista¹, mládež², dítě, mladistvý a k tomu nějaké doplňující informace, jako např. zda to bylo pod vlivem alkoholu, se zbraní apod. Ukázka tabulky poskytnuté Policií ČR je uvedena na Obrázku 2.

Činy spáchané ve sledovaných místech za období od 1.1.2008 do 31.12.2008

Kraj: **02 - KŘP JIHOČESKÉHO KRAJE**
 Okres: **05 - ÚO PÍSEK**
 Útvar: **15 - OOP PÍSEK**

Sledované údaje	Celková	Vraždy	Loupeže	Ublíž.	Rvačky	Znásil.	Vl. obch.	Vl. byty	Krád.aut	Podv.
	101	101	131	151	161	201	311	371	431	511
	902	106	132							
Zjištěno	679	1	12	11	0	2	12	8	45	10
Objasněno	295	1	9	9	0	1	2	1	9	6
Tj. %	43,45	100,00	75,00	81,82	0,00	50,00	16,67	12,50	20,00	60,00
Dodatečně objasněno	12	0	2	0	0	0	0	0	1	0
Objasněno celkem	307	1	11	9	0	1	2	1	10	6
Spácháno na ulici	315	1	9	6	0	2	7	2	42	5
Objasněno (ulice)	119	1	7	4	0	1	1	0	8	2
Tj. %	37,78	100,00	77,78	66,67	0,00	50,00	14,29	0,00	19,05	40,00
Dodat. obj. (ulice)	1	0	0	0	0	0	0	0	1	0
Objasněno celkem (ulice)	120	1	7	4	0	1	1	0	9	2
Hodina spáchání - 01 až 06	18	0	2	0	0	1	1	0	0	0
ve sled. místě - 07 až 12	41	0	0	1	0	0	0	0	0	0
- 13 až 18	58	1	3	1	0	0	0	0	2	0
- 19 až 24	42	0	0	2	0	0	2	0	1	0
- nezjištěno	520	0	7	7	0	1	9	8	42	10
Den spáchání - pondělí	41	0	3	2	0	0	0	1	8	0

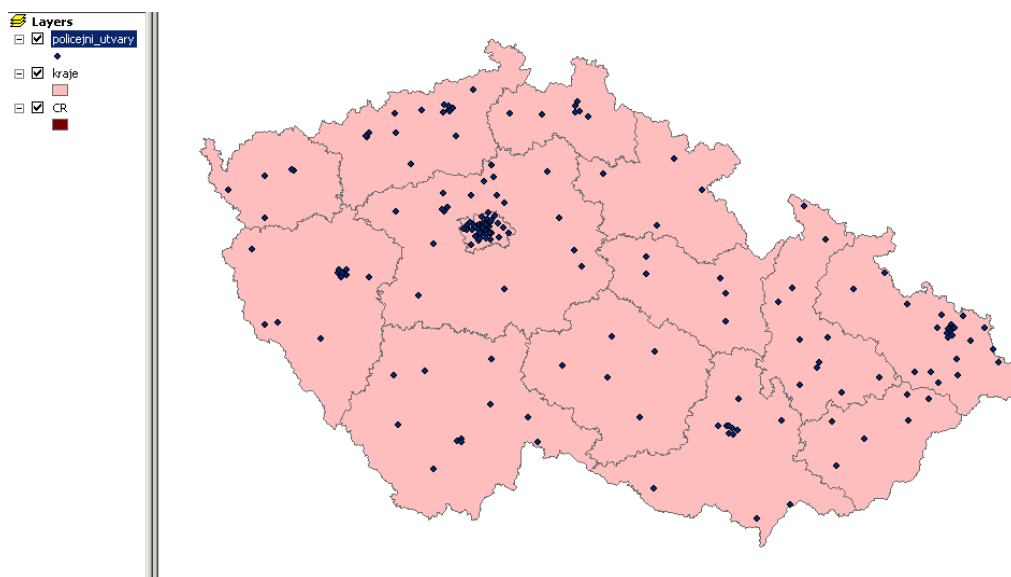
Obr. 2: Ukázka údajů o trestných činech v roce 2008

Policie ČR poskytla pro tuto práci tematické neprostorové informace. K těmto informacím bylo potřeba dodat ještě prostorovou informaci, souřadnice v S-JTSK jednotlivých policejních útvarů, ve kterých byla zjištěna trestná činnost. Tyto souřadnice byly převzaty z Územně identifikačního registru ÚIR-ZJS, což je soustava databázových číselníků jednotek územně správního, technického a sídelního členění státu, poskytovaná na internetových stránkách Českého statistického úřadu[5].

¹ Recidivista- občan, který spáchá trestný čin opakovaně. Může se jednat i o dítě a mladistvého.

² Mládež – děti do 15 let a zároveň mladiství od 15-18 let. Pokud je zaznamenán trestný čin dítěte nebo mladistvého, je automaticky započten i do trestných činů mládeže.

Takto upravená data o policejních útvarech, byla nahrána do prostředí ArcGIS 9.3, kde byla následně převedena do formátu shapefile. Pro účely této práce bylo nutné ještě získat geometrii celé České republiky a jednotlivých krajů. Ty byly získány z digitální vektorové geografické databáze pro území České republiky, ArcČR 500[16]. Tímto byla získána prostorová data, se kterými bylo možné dále pracovat. Na Obrázku 6 jsou znázorněny kraje ČR a policejní útvary, ve kterých byla zjištěna trestná činnost.



Obr. 3: Kraje a policejní útvary ve formátu shapefile

2.3 Příprava dat – ETL proces

Příprava dat je velmi důležitý krok. První částí přípravy dat je tzv. ETL proces (Extraction, Transformation, Load proces), v němž se jedná o extrakci dat ze zdrojových systémů, jejich transformaci do vhodné podoby a následné uložení dat do databáze.

2.3.1 Extrakce

Pro účely této práce byly zvoleny záznamy o loupežích a to v roce 2008, aby data byla co nejaktuálnější. Loupeže byly zvoleny ze dvou důvodů. Prvním důvodem bylo to, že pro ně byly zaznamenány údaje v mnoha policejních útvech. Druhým důvodem bylo, že tento trestný čin má jako jeden z mála činů mnoho dětských a mladistvých pachatelů.

Protože pro aplikaci na konkrétních datech byla zvolena metoda asociace, byly do atributové tabulky policejních útvarů v prostředí ArcGIS 9.3 přidány tyto sloupce s neprostorovými informacemi:

- loupez_pondeli
- loupez_utery
- loupez_streda
- loupez_ctvrtek
- loupez_patek
- loupez_sobota
- loupez_nedele
- loupez_01_06
- loupez_07_12
- loupez_13_18
- loupez_19_24
- loupez_recidivista
- loupez_deti
- loupez_mladez
- loupez_mladistvi
- loupez_mladistvi_alkohol

Buňky v přidávaných sloupcích mohly nabývat pouze hodnot 0 nebo 1. Pokud byla v policejním útvaru spáchána alespoň jedna loupež v některém dni v týdnu, v některém z časových rozmezí a nebo některým z pachatelů, poté se v konkrétní buňce daného sloupce objevila hodnota 1. Pokud se v daný den nebo v daném časovém úseku loupež nestala, buňka obsahovala hodnotu 0.

Výsledkem extrakce dat byla tedy atributová tabulka, ve které byl uveden název policejního útvaru, souřadnice v S-JTSK a dále všechny výše uvedené sloupce.

2.3.2 Transformace

Po extrahování potřebných údajů z poskytnutých dat, bylo nutné prostorovou informaci, geometrii uloženou ve formátu shapefile, převést do formátu, ve kterém ukládá prostorová data databázový systém Oracle, tedy SDO_GEOMETRY.

Dle [17] je pomocí objektového typu SDO_GEOMETRY uložen geometrický popis prostorového objektu. Libovolná tabulka, která obsahuje sloupec tohoto typu, musí mít jiný sloupec, který definuje unikátní primární klíč pro tuto tabulku. Objektový typ SDO_GEOMETRY je v prostředí Oracle definován následujícím způsobem.

```
CREATE TYPE sdo_geometry AS OBJECT (  
    Sdo_gtype          NUMBER  
    Sdo_srid           NUMBER  
    Sdo_point          SDO_POINT_TYPE  
    Sdo_elem_info      MDSYS.SDO_ELEM_INFO_ARRAY  
    Sdo_ordinates      MDSYS.SDO_ORDINATE_ARRAY  
);
```

Objektový typ SDO_GEOMETRY zahrnuje tyto atributy:

- `sdo_gtype` – geometrický typ geometrického popisu prvku,
- `sdo_srid` – identifikace souřadnicového systému pro geometrický popis prostorového prvku,
- `sdo_point` – doporučuje se ukládat do něj souřadnice bodů
- `sdo_elem_info` – popisuje způsob interpretace souřadnic geometrického popisu prvků uložených v atributu `sdo_ordinates`,

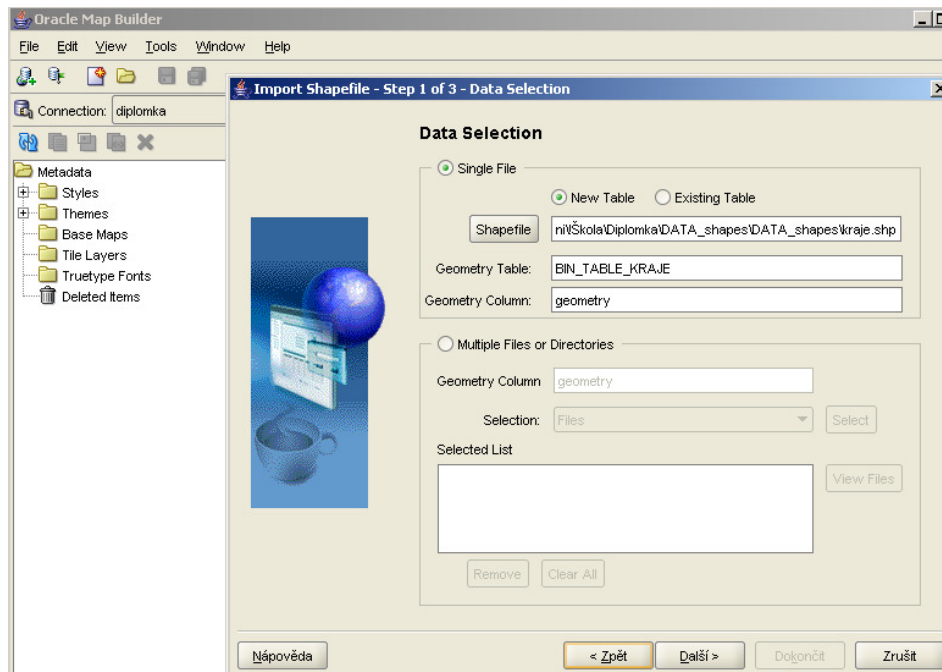
- `sdo_ordinates` – zde jsou uloženy souřadnice bodů, které vytváří geometrický popis hranice prostorového prvku[17].

Aby mohly být vyhledány regionální vzory, bylo nutné do tohoto formátu převést geometrii celé ČR, geometrii krajů ČR a v neposlední řadě geometrii policejních útvarů.

Aby bylo možné toto splnit, musela být vytvořena databáze, do níž se tato data měla transformovat. Databáze byla vytvořena v grafickém nástroji databázového systému Oracle, SQL Developer.

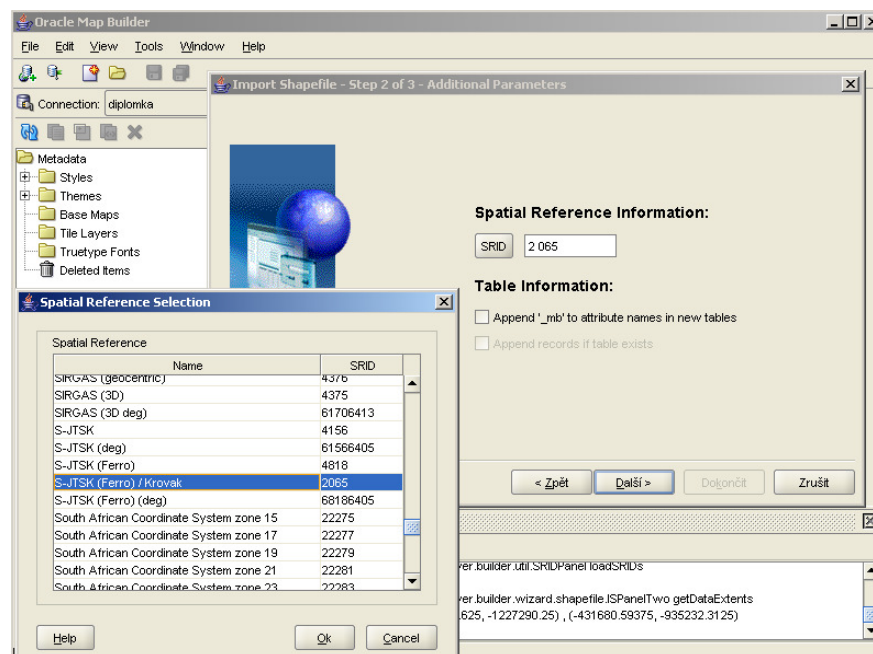
Samotný převod shapefile do `SDO_GEOMETRY` byl také proveden pomocí jedné z aplikací, kterou poskytuje databázový systém Oracle, konkrétně aplikací Oracle Map Builder[18]. Tato aplikace umožňuje vytvářet, spravovat a nahlížet na metadata a prostorovou informaci. Umožňuje také právě import rastrů, shapefilů a metadat do databáze Oracle.

Prvním krokem použití této aplikace bylo připojení dříve vytvořené databáze. Po připojení databáze, která nemusela obsahovat žádné tabulky, bylo možné spustit funkci *Import Shapefile* v záložce *Tools*. Následně bylo zobrazeno okno, ve kterém byl zvolen shapefile, který měl být převeden do databáze. V dalším kroku, bylo nutné zadat, zda se mají data ukládat do nové či stávající tabulky v databázi, jak se tato tabulka bude jmenovat, či jak se jmenuje a do jakého sloupce se má uložit geometrie. Toto výběrové okno je zobrazeno na Obrázku 4.



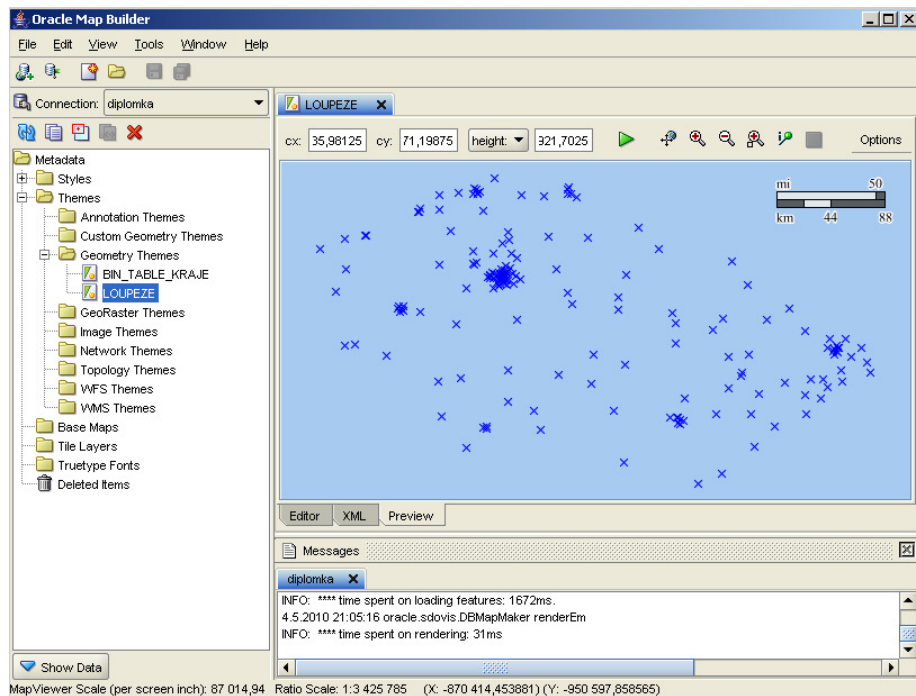
Obr. 4: Okno pro výběr shapefile importovaného do databáze Oracle

Po zadání všech požadovaných informací bylo v aplikaci Oracle Map Builder zobrazeno další výběrové okno, ve kterém bylo možné zvolit souřadnicový systém, ve kterém se má geometrie převést. Jak je vidět na Obrázku 5, byl vybrán systém S-JTSK.



Obr. 5: Volba souřadnicového systému pro převod do SDO_GEOMETRY

V dalším kroku této transformace bylo možné založit tzv. *Thema*, které umožňuje podívat se na převedená data a zkontrolovat, jestli je geometrie uložena správně. Toto *Thema* převedené tabulky policejních útvarů je zachyceno na Obrázku 6.



Obr. 6: Thema LOUPEZE získané převodem geometrie policejních útvarů

Takto byla zkontrolována geometrie celé České republiky, krajů a jednotlivých policejních útvarů. Vše bylo v pořádku a tak mohla být funkce *Import Shapefile* dokončena.

2.3.3 Načtení do databáze

Ukončením funkce *Import Shapefile* byly v databázi vytvořeny tři nové tabulky, *BIN_TABLE_CR*, *BIN_TABLE_KRAJE* a *LOUPEZE*. Všechny tabulky obsahují mimo jiné sloupec *ID* a jeden sloupec nazvaný *GEOMETRY*, v němž je uložena geometrie ve formátu *SDO_GEOMETRY*. Tabulka *LOUPEZE* má navíc sloupce s neprostorovými informacemi, uvedené v kapitole 2.3.1.

Ukázka tabulky BIN_TABLE_KRAJE je uvedena na následujícím obrázku.

	PERI...	NAZEV	OB91	OB01	OB_311202	NUTS3	NK	KN	N...	NAZEV_ENG	ID	GEOM
1..	419095	Liberecký	425120	430820	427321	CZ051	LB	07	CZ05	Liberecky	1	MDSY
2..	601702	Ústecký	824461	826869	819712	CZ042	US	06	CZ04	Ustecky	2	MDSY
3..	150985	Hlavní město Praha	1214174	1180131	1161938	CZ011	HP	01	CZ01	Hlavni mesto Praha	3	MDSY
4..	414085	Karlovarský	301985	306887	304220	CZ041	KA	05	CZ04	Karlovarsky	4	MDSY
5..	554071	Královéhradecký	552809	554594	548437	CZ052	KH	08	CZ05	Kralovehradecky	5	MDSY
6..	1019320	Středočeský	1112882	1128766	1128674	CZ021	ST	02	CZ02	Stredocesky	6	MDSY
7..	582039	Píseňský	558307	553848	549374	CZ032	PL	04	CZ03	Pizensky	7	MDSY
8..	495415	Pardubický	508718	510126	506534	CZ053	PA	09	CZ05	Pardubicky	8	MDSY
9..	643412	Olomoucký	642796	642346	636750	CZ071	OL	12	CZ07	Olomoucky	9	MDSY
10..	541629	Moravskoslezský	1283271	1277412	1262660	CZ081	MS	13	CZ08	Moravskoslezsky	10	MDSY
11..	577442	Vysočina	521068	521308	517630	CZ061	VY	10	CZ06	Vysocina	11	MDSY
12..	765455	Jihočeský	622889	630190	625097	CZ031	JC	03	CZ03	Jihocesky	12	MDSY
13..	386819	Zlínský	596903	597848	593130	CZ072	ZL	14	CZ07	Zlinsky	13	MDSY
14..	736104	Jihomoravský	1136832	1134170	1121792	CZ062	JM	11	CZ06	Jihomoravsky	14	MDSY

Obr. 7: Tabulka BIN_TABLE_KRAJE a její atributy

2.4 Příprava dat pro prostorové dolování dat

ETL proces sám o sobě však připravuje data pouze pro dolování dat. V každé tabulce, BIN_TABLE_CR, BIN_TABLE_KRAJE i LOUPEZE je definována geometrie každé položky. Tyto tabulky bylo nutné „spojit dohromady“, a to použitím funkce prostorového dolování dat *Spatial Binning*. Aby tato funkce mohla být použita, bylo nutné do tabulek BIN_TABLE_CR, BIN_TABLE_KRAJE přidat sloupec s názvem BIN a do tabulky LOUPEZE přidat sloupec s názvem ID_BIN.

Nad tabulkami BIN_TABLE_CR a BIN_TABLE_KRAJE byla nejprve spuštěna procedura SDO_SAM.BIN_GEOMETRY, jejíž parametry musí odpovídat níže uvedenému formátu. Tato procedura rozdělila policejní útvary podle jejich prostorové informace do jednotlivých binů, vytvořených podle geografických areálů.

Formát:

```
SDO_SAM.BIN_GEOMETRY (  
Geom                IN SDO_GEOMETRY  
Tol                 IN SDO_DIM_ARRAY  
Bin_tablename       IN VARCHAR2  
Bin_colname         IN VARCHAR2  
) RETURN NUMBER;
```

Parametr geom identifikuje geometrii, která má být rozdělena biny. Parametr tol definuje hodnotu tolerance, která udává, do jaké vzdálenosti jsou dva velmi blízké objekty brány jako jeden. Parametr bin_tablename udává jméno tabulky, která uchovává geometrii, která má být rozdělena do binů a parametr bin_colname je přímo název sloupce uchovávajícího geometrii.

V našem konkrétním případě byl nejprve vytvořen pouze jeden bin podle geometrie České republiky, přičemž se v tabulce BIN_TABLE_CR do sloupce BIN uložil jeho identifikátor. Následně bylo vytvořeno 14 binů podle jednotlivých krajů a v tabulce BIN_TABLE_KRAJE byly také do sloupce BIN uloženy jejich identifikátory.

Spojení tabulky (*Spatial Binning*) LOUPEZE s tabulkami BIN_TABLE_CR a BIN_TABLE_KRAJE bylo provedeno pomocí procedury SDO_SAM.BIN_LAYER, která má následující parametry.

```
SDO_SAM.BIN_LAYER (  
tablename           IN VARCHAR2  
colname             IN VARCHAR2  
bin_tablename       IN VARCHAR2  
bin_colname         IN VARCHAR2  
bin_id_colname      IN VARCHAR2  
commit_interval     IN NUMBER DEFAULT 20  
) ;
```


Parametr `tablename` identifikuje tabulku dolování dat (v našem konkrétním případě tabulka `LOUPEZE`). Parametr `colname` identifikuje sloupec v `tablename`, který uchovává geometrii (sloupec `GEOMETRY`). `Bin_tablename` je název tabulky, která uchovává informaci o prostorových binech (`BIN_TABLE_CR`, `BIN_TABLE_KRAJE`), a `bin_colname` je název sloupce v `bin_tablename` (`BIN`). Parametr `bin_id_colname` je sloupec v tabulce `tablename`. Do tohoto sloupce se uloží identifikátor binu, do něhož spadá policejní útvar (sloupec `ID_BIN`). Poslední parametr `commit_interval` vyjadřuje po kolika provedených operacích maximálně, mají být identifikátory binů uloženy do tabulky dolování dat.

Teprve po provedení těchto operací byla získána materializovaná data, tzn. data zahrnující prostorovou informaci, připravená pro prostorové dolování dat. Z tabulky `LOUPEZE` bylo následně vytvořeno 14 tabulek, rozdělujících policejní útvary podle jednotlivých krajů.

Bylo však nutné si uvědomit, že do každé metody prostorového dolování dat vstupují data v jiném formátu. Pro zvolenou metodu asociace, bylo třeba tabulky vytvořené podle jednotlivých krajů upravit do formátu tzv. transakčních dat, popsanych v mé diplomové práci v kapitole 3.2.2. V praxi to znamenalo vytvořit těchto 14 tabulek:

- `TRANSAKCNI_DATA_JIHOCESKY`
- `TRANSAKCNI_DATA_PLZENSKY`
- `TRANSAKCNI_DATA_KARLOVARSKY`
- `TRANSAKCNI_DATA_USTECKY`
- `TRANSAKCNI_DATA_LIBERECKY`
- `TRANSAKCNI_DATA_KRALOVEHRADECKY`
- `TRANSAKCNI_DATA_PARDUBICKY`
- `TRANSAKCNI_DATA_STREDOCESKY`
- `TRANSAKCNI_DATA_HLAVNI_MESTO_PRAHA`
- `TRANSAKCNI_DATA_VYSOCINA`
- `TRANSAKCNI_DATA_OLOMOUCKY`
- `TRANSAKCNI_DATA_MORAVSKOSLEZSKY`

- TRANSAKCNI_DATA_ZLINSKY
- TRANSAKCNI_DATA_JIHOMORAVSKY

Do těchto tabulek transakčních dat, rozdělených podle krajů, nebylo již nutné uvádět mnoho atributů. Metodě asociace stačí dva identifikátory, ostatní jsou již nadbytečné. Proto tabulky transakčních dat jednotlivých krajů mohly obsahovat pouze tyto sloupce:

- KOD_CINU – může nabývat hodnot (loupez_pondeli, loupez_utery, loupez_streda, loupez_ctvrtek, loupez_patek, loupez_sobota, loupez_nedele, loupez_01_06, loupez_07_12, loupez_13_18, loupez_19_24, loupez_recidivista, loupez_dite, loupez_mladez, loupez_mladistvi, loupez_mladistvi_alkohol)
- NAZEV_UTVARU – může nabývat hodnot názvů všech policejních útvarů. K řádkům ve sloupci KOD_CINU však přiřazuje policejní útvary, v nichž se pro daný kód, loupež skutečně stala.
- ID_UTVARU – identifikátor policejního útvaru, kde se stala loupež

Obrázek 8 zobrazuje ukázkou tabulky TRANSAKCNI_DATA_PLZENSKY, tedy transakční data pro Plzeňský kraj.

ID	KOD_CINU	NAZEV_UTVARU	ID_UTVARU
1	59 loupez_pondeli	Domazlice	74
2	63 loupez_pondeli	Plzen4	80
3	64 loupez_pondeli	Plzen2	81
4	65 loupez_pondeli	Plzen1	82
5	66 loupez_pondeli	Plzen_Bory	84
6	67 loupez_pondeli	Plzen_Vinice	85
7	68 loupez_pondeli	Plzen	86
8	69 loupez_pondeli	Rokycany	87
9	70 loupez_pondeli	Chodov	88
10	72 loupez_pondeli	Tachov	90
11	197 loupez_utery	Domazlice	74
12	200 loupez_utery	Klatovy	79
13	201 loupez_utery	Plzen4	80
14	202 loupez_utery	Plzen2	81
15	203 loupez_utery	Plzen1	82
16	204 loupez_utery	Plzen_Skvrmany	83
17	205 loupez_utery	Plzen_Bory	84
18	206 loupez_utery	Plzen_Vinice	85
19	207 loupez_utery	Plzen	86
20	208 loupez_utery	Rokycany	87
21	323 loupez_streda	Domazlice	74
22	328 loupez_streda	Klatovy	79

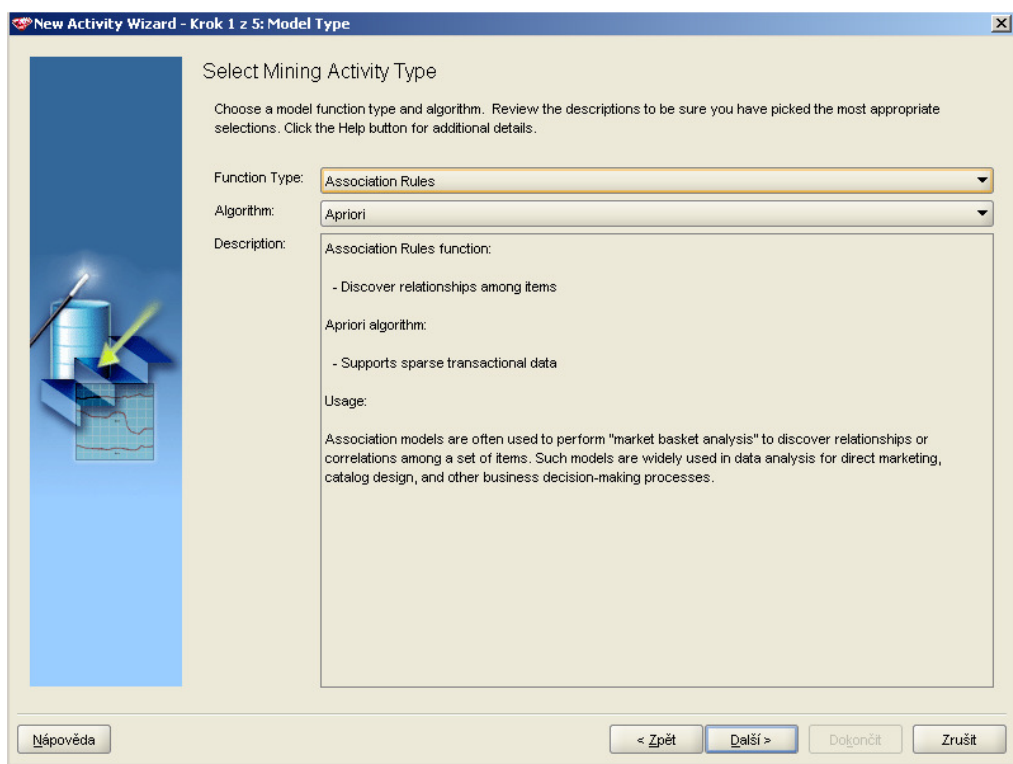
Obr. 8: Tabulka transakčních dat pro Plzeňský kraj

2.5 Aplikace metody vyhledávání asociačních pravidel

Na data upravená do formátu transakčních dat už mohla být aplikována metoda asociace, která měla odhalit vztahy a závislosti vyjádřené pomocí asociačních pravidel. Pro jejich vyhledání bylo využito uživatelské rozhraní Oracle Data Miner, které bylo společně s algoritmy, které podporuje, více přiblíženo v mé diplomové práci v kapitole 4.1.

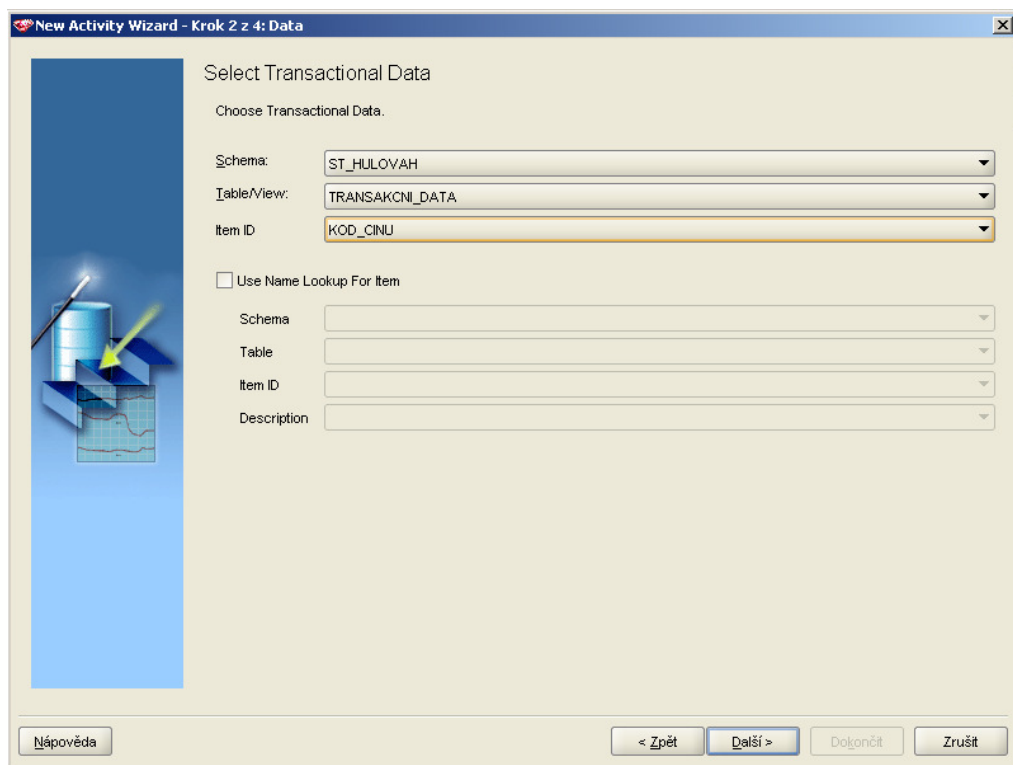
2.5.1 Hledání regionálních vzorů

Po spuštění prostředí Oracle Data Miner bylo opět v první řadě nutné připojit stávající databázi Oracle. Vlastní aplikace potom začala tím, že ze záložek na horním panelu byla vybrána záložka *Activity* a v rozbalovacím menu byla zvolena funkce *Build*. Následně bylo zobrazeno okno, které sloužilo pro výběr metody a algoritmu, jak je vidět na Obrázku 9.



Obr. 9: Výběr metody prostorového dolování dat

V tomto okně byl nastaven typ metody, tedy vyhledávání asociačních pravidel a dále algoritmus *APRIORI*, kterým se asociační pravidla vyhledají. V okně se zobrazuje i textový popis dané metody a algoritmu. Po vyplnění těchto údajů bylo zobrazeno další okno, které sloužilo pro volbu tabulky. Zde byly zvoleny postupně všechny tabulky transakčních dat, pro něž se měla asociační pravidla vyhledat. Dále zde bylo nutné nastavit identifikátor transakce. Jako identifikátor byl vybrán atribut KOD_CINU, protože právě tento atribut byl ten, pro který byly vyplňovány údaje v transakčních tabulkách. Toto okno je zobrazeno na následujícím obrázku.



Obr. 10: Výběr tabulky transakčních dat a identifikátoru transakce

Po vyplnění této části bylo zobrazeno další výběrové okno, ve kterém byly požadovány další identifikátory transakce, což v datech o loupežích byly ID_UTVARU a NAZEV_UTVARU. Poté bylo třeba nastavit název nové tabulky, do níž se vygenerují asociační pravidla.

Asociačními pravidly jsou charakterizovány vztahy mezi jednotlivými položkami. Obecně jsou tato pravidla generována v tomto tvaru: *Pokud je v souboru dat obsažena položka a a zároveň položka b, pak platí pravidlo s.* Konkrétní případ vygenerovaného asociačního pravidla je uveden na Obrázku 11, kde je vidět, že pokud se loupež stala v pondělí v rozmezí od 13 do 18 hodin, pak jí ve 45% případů spáchalo dítě.

```
Rule Detail
IF
KOD_CINU.loupeze_13_18= 1 AND KOD_CINU.loupez_pondeli= 1

THEN
KOD_CINU.loupeze_deti= 1

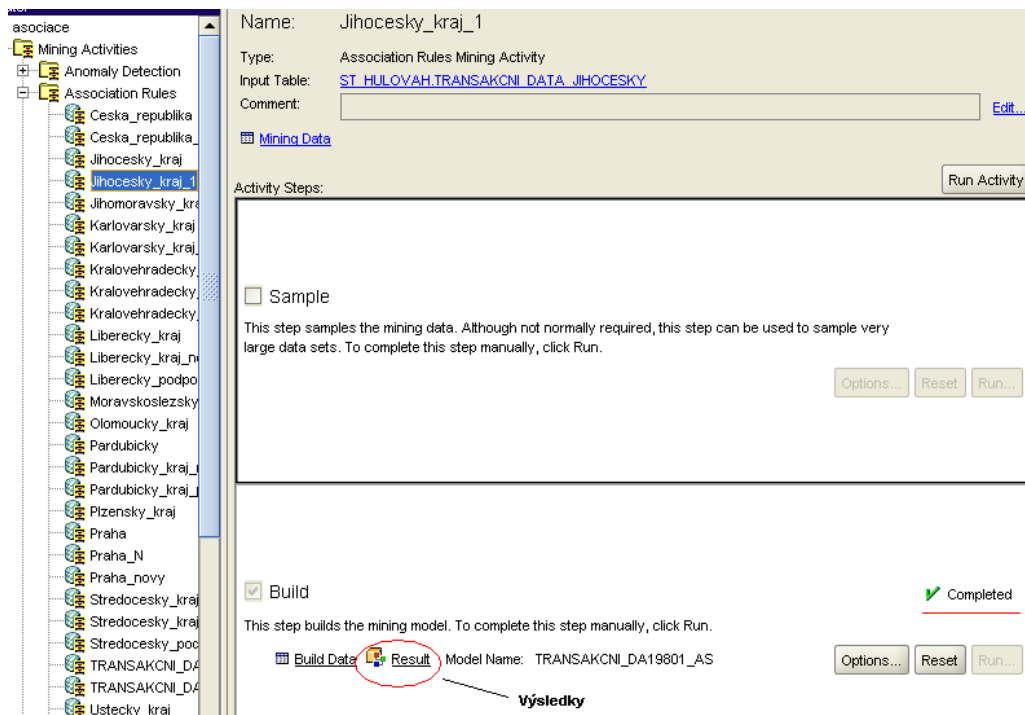
Support (%)= 100%
Confidence (%)= 45,45%
```

Obr. 11: Výpis asociačního pravidla

Na obrázku je vidět, že parametr *Podpora* pro tohle pravidlo je 100% a parametr *Spolehlivost* je 45%. Toto pravidlo už má určitou váhu, v podstatě říká, že se ve 100% případů stala loupež v pondělí od 13 do 18 hodin a ve 45% ze všech těchto případů ji spáchalo dítě.

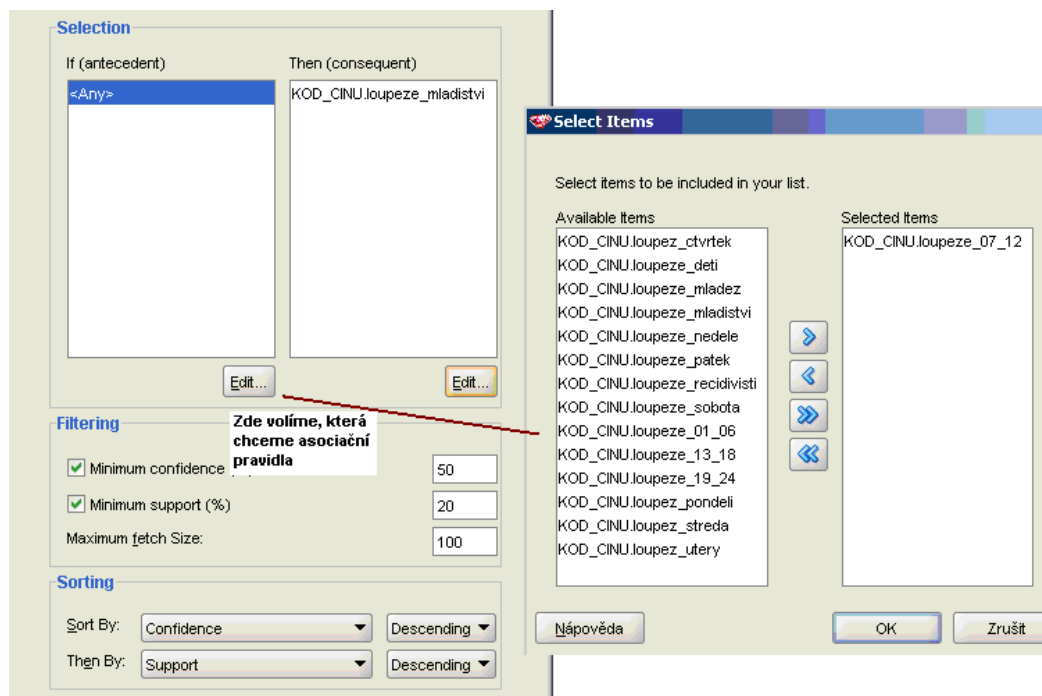
Před samotným vygenerováním pravidel bylo nutné zvážit, jakou spodní hranici nastavit pro parametry *Podpora* a *Spolehlivost*, tak aby pravidlo mělo nějaký význam a mohlo být považováno za relevantní. Parametr *Podpora* byl nastaven pro všechny kraje na 30% a parametr *Spolehlivost* na 20%. To znamená, že minimálně ve 30% případů ze všech možných se musely dvě položky vyskytovat společně. Spolehlivost pak vyjadřuje, že minimálně ve 20% případů z těchto vybraných, se musela vyskytovat další položka, aby bylo vygenerováno pravidlo.

Tyto hodnoty byly zadány ve výběrovém okně jako poslední parametry a poté již bylo vše připraveno pro vygenerování konkrétních asociačních pravidel. Po spuštění vyhledávání bylo zobrazeno opět základní okno, uvedené na následujícím obrázku.



Obr. 12: Vyhledání výsledků metody asociace

V pravé dolní části bylo možné sledovat průběh vyhledávání a zde je již vidět, že výsledky jsou kompletní. Pro zobrazení výsledků byla vybrána možnost *Result*. Po jejím zvolení bylo zobrazeno okno, v němž bylo možné vybrat, jaká asociční pravidla má program vypsát. Toto okno je zobrazeno na Obrázku 13.



Obr. 13: Tabulka pro výběr asocičních pravidel

V tomto okně bylo možné nastavit jak činy, které by se měly společně vyskytovat, tak také kterého činu se má vygenerované pravidlo týkat. Protože byl na začátku vymezen problém kriminality dětí a mladistvých, byly v tomto okně postupně vybrány v druhém sloupci děti a mladiství. Pokud tedy byly v prvním sloupci ponechány všechny možnosti a ve druhém sloupci byli vybráni mladiství, zobrazila se pravidla týkající se pouze nich. Ve spodní části této tabulky bylo také možné nastavit minimální hodnoty atributů pro parametry *Podpora* a *Spolehlivost*, pro něž se měla pravidla vypsát. Výpis asociačních pravidel je uveden na Obrázku 14.

Rule Id	If (condition)	Then (association)	Support (%)
496	KOD_CINU.loupez_streda= 1 AND KOD_CINU.loupez_mladez= 1	KOD_CINU.loupez_mladistvi= 1	100,0000
613	KOD_CINU.loupez_19_24= 1 AND KOD_CINU.loupez_mladez= 1	KOD_CINU.loupez_mladistvi= 1	100,0000
484	KOD_CINU.loupez_streda= 1 AND KOD_CINU.loupez_pondeli= 1	KOD_CINU.loupez_mladistvi= 1	100,0000
526	KOD_CINU.loupez_01_06= 1 AND KOD_CINU.loupez_mladez= 1	KOD_CINU.loupez_mladistvi= 1	100,0000
604	KOD_CINU.loupez_19_24= 1 AND KOD_CINU.loupez_pondeli= 1	KOD_CINU.loupez_mladistvi= 1	100,0000
679	KOD_CINU.loupez_pondeli= 1 AND KOD_CINU.loupez_nedele= 1	KOD_CINU.loupez_mladistvi= 1	100,0000
508	KOD_CINU.loupez_streda= 1 AND KOD_CINU.loupez_sobota= 1	KOD_CINU.loupez_mladistvi= 1	100,0000
622	KOD_CINU.loupez_19_24= 1 AND KOD_CINU.loupez_sobota= 1	KOD_CINU.loupez_mladistvi= 1	100,0000
691	KOD_CINU.loupez_mladez= 1 AND KOD_CINU.loupez_nedele= 1	KOD_CINU.loupez_mladistvi= 1	100,0000
334	KOD_CINU.loupez_deti= 1 AND KOD_CINU.loupez_streda= 1	KOD_CINU.loupez_mladistvi= 1	100,0000
704	KOD_CINU.loupez_mladez= 1 AND KOD_CINU.loupez_recidivistvi= 1	KOD_CINU.loupez_mladistvi= 1	85,7143

Rule Detail

IF
KOD_CINU.loupez_streda= 1 AND KOD_CINU.loupez_mladez= 1

THEN
KOD_CINU.loupez_mladistvi= 1

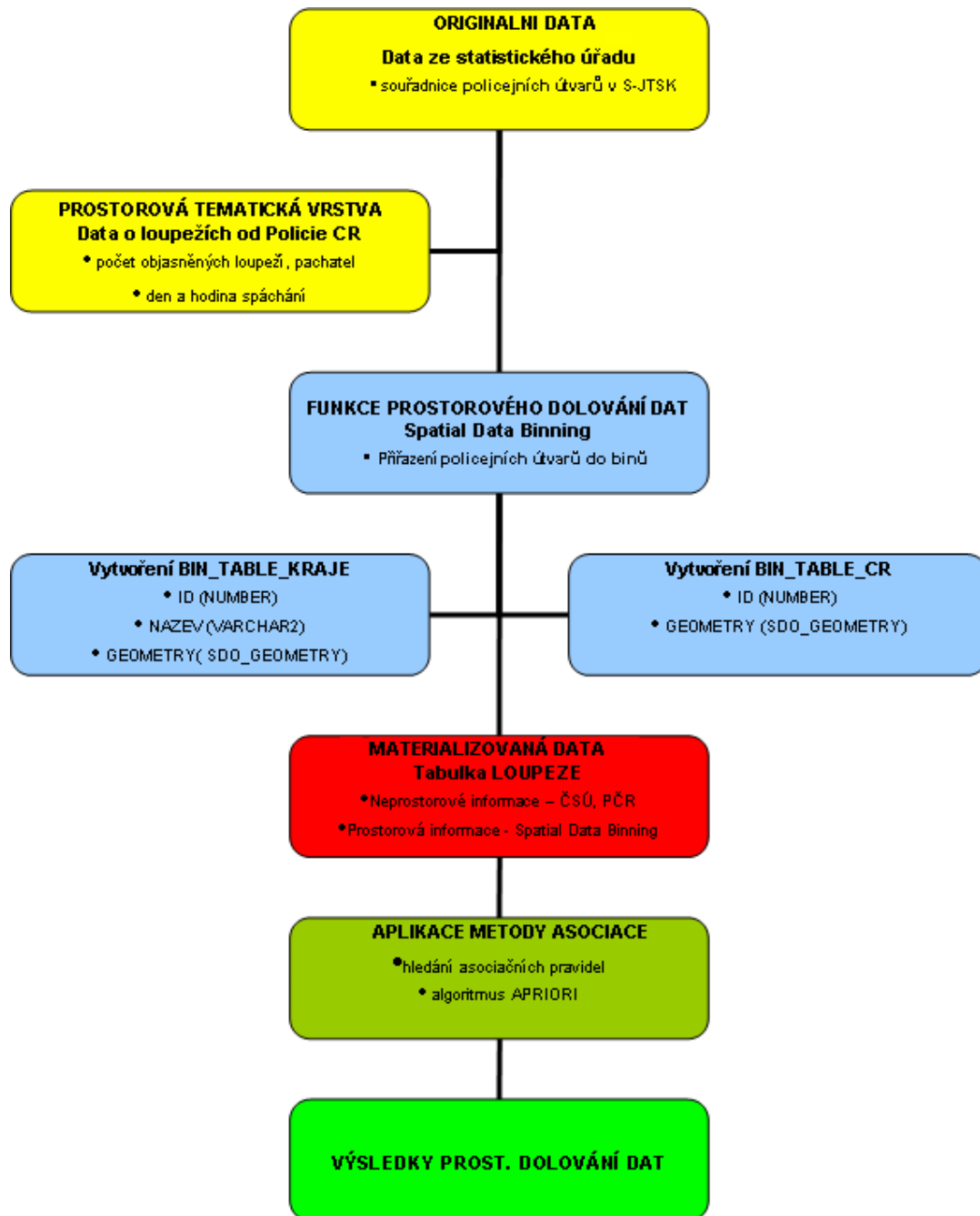
Support (%)= 100
Confidence (%)= 54,55

Obr. 14: Asociační pravidla pro kriminalitu mladistvých

Tímto způsobem byla pravidla vygenerována pro každý kraj i plošně přes celou Českou republiku. V reálném případě by takto zpracovaná data dostal k vyhodnocení analytik, který by ve výsledcích viděl víc než „obyčejný uživatel“, ale v rámci této práce bylo nutné data vyhodnotit a porovnat laickým pohledem.

O výsledky této práce projevil zájem kpt. Mgr. Jan Melša, Preventivně informační odbor Policejního prezidia České republiky.

Jednotlivé kroky procesu aplikace metody asociace, které byly podrobně popsány v této kapitole, jsou shrnuty na následujícím Obrázku 15, který je obdobou Obr. 1, uvedeného v kapitole 1.2, avšak zde jsou již konkrétně uvedeny funkce, metody a tabulky, které byly v této práci v jednotlivých fázích procesu prostorového dolování dat vytvořeny a použity. Jednotlivé kroky procesu jsou do sebe barevně odlišeny.



Obr. 15: Fáze procesu prostorového dolování dat s reálnými daty

3 Porovnání a zhodnocení výsledků

V následující kapitole jsou uvedeny tabulky, v nichž jsou vypsány výsledky asociačních pravidel pro každý kraj. Pro přehlednost je vždy uvedeno maximálně 10 nejvýznamnějších pravidel. Nejvýznamnějšími pravidly jsou myšlena ta pravidla, která mají právě nejvyšší hodnoty parametrů *Podpora* a *Spolehlivost*.

3.1 Loupeže spáchané dětmi v krajích ČR

U jednotlivých tabulek jsou uvedeny vždy hodnoty parametrů *Podpora* a *Spolehlivost*, vyjadřující procento společného výskytu dvou položek, týkajících se dne a času spáchání. V tabulce může být případně uvedeno, pokud byl Policií ČR zaevidován jako pachatel mládež či recidivista, tak v kolika procentech z těchto případů se jednalo o dětského pachatele.

Tab. 2: Asociační pravidla pro Jihočeský kraj

ID	Jihočeský kraj	podpora	spolehlivost
1	loupež v pondělí od 13 do 18 hod.	100%	45%
2	loupež spáchaná mládeží od 13 do 18 hod.	100%	45%
3	loupež spáchaná mládeží v úterý	83%	45%
4	loupež spáchaná mládeží v pondělí	83%	45%
5	loupež spáchaná recidivistou v pondělí	83%	45%
6	loupež spáchaná recidivistou od 13 do 18 hod.	83%	45%

Z tabulky pravidel pro Jihočeský kraj je vidět, že loupež byla ve 100% případů spáchaná v pondělí od 13 do 18 hodin, a z toho ve 45% případů ji spáchalo dítě. Pokud byla loupež spáchaná od 13 do 18 hodin a jako pachatel byla evidována mládež, pak se také ve 45% případů jednalo o dětského pachatele. Takto lze interpretovat i ostatní pravidla.

Tab. 3: Asociační pravidla pro Plzeňský kraj

ID	Plzeňský kraj	podpora	spolehlivost
1	loupež spáchána mládeží od 19 do 24 hod.	80%	33%
2	loupež spáchána mládeží v pondělí	66%	33%
3	loupež spáchána mládeží ve čtvrtek	66%	33%

V Plzeňském kraji zřejmě není kriminalita dětí příliš velká, z uvedených pravidel se dá pouze říci, že v 80% případů byla loupež spáchána mládeží od 19 do 24 hodin a z toho pouze ve 33% případů bylo pachatelem dítě.

Tab. 4: Asociační pravidla pro Karlovarský kraj

ID	Karlovarský kraj	podpora	spolehlivost
1	loupež v úterý od 13 do 18 hod.	33%	20%
2	loupež spáchána mládeží v úterý	33%	20%
3	loupež spáchána mládeží v pondělí	33%	20%

Hodnoty ve výše uvedené tabulce pravidel pro Karlovarský kraj se dají považovat, co se týče kriminality dětí, téměř za bezvýznamné. Podpora i spolehlivost jsou velmi nízké hodnoty, přesto tato pravidla měla nejvyšší hodnoty ze všech vygenerovaných pravidel, proto jsou zde uvedena.

Tab. 5: Asociační pravidla pro Ústecký kraj

ID	Ústecký kraj	podpora	spolehlivost
	loupež v pondělí od 1 do 6 hod.	81%	50%
	loupež ve středu od 1 do 6 hod.	81%	50%
	loupež spáchána mládeží v pondělí	81%	50%
	loupež spáchána mládeží od 1 do 6 hod.	81%	50%
	loupež v sobotu od 1 do 6 hod.	80%	44%
	loupež v neděli od 1 do 6 hod.	80%	44%
	loupež spáchána recidivistou od 1 do 6 hod.	80%	44%

V sousedním kraji „téměř bezkriminalitového“ Karlovarského kraje je situace docela jiná. V tabulce je vidět, že zde děti páchají loupeže hlavně v brzkých ranních hodinách, což se liší od krajů, jejichž pravidla byla zatím uvedena, v nichž byly trestné činy spáchány hlavně od 13 do 18 hodin. Nejvíce loupeží bylo dětmi spácháno v pondělí a ve středu od 1 do 6 hodin ráno, a to v rovných 50% případů.

Tab. 6: Asociační pravidla pro Liberecký kraj

ID	Liberecký kraj	podpora	spolehlivost
1	loupež ve čtvrtek od 7 do 12 hod.	50%	11%
2	loupež ve středu od 7 do 12 hod.	50%	11%
3	loupež ve středu od 1 do 6 hod.	50%	11%

Liberecký kraj viditelně spadá také spíše do skupiny těch s nízkou kriminalitou dětí. V tabulce jsou uvedena pouze tři pravidla, která se opět dají považovat téměř za bezvýznamná, děti spáchaly loupež pouze v 11% případů.

Tab. 7: Asociační pravidla pro Královeshradecký kraj

ID	Královeshradecký kraj	podpora	spolehlivost
1	loupež ve středu od 1 do 6 hod.	100%	25%
2	loupež spáchána mládeží od 1 do 6 hod.	100%	25%
3	loupež spáchána mládeží v neděli	100%	25%
4	loupež v sobotu od 7 do 12 hod.	100%	25%
5	loupež v neděli od 7 do 12 hod.	100%	25%
6	loupež v pondělí od 7 do 12 hod.	100%	25%
7	loupež ve čtvrtek od 7 do 12 hod.	100%	25%

Tabulka zobrazuje vždy dvě položky, které se staly společně ve 100% případů, avšak pouze čtvrtinu z nich spáchaly děti, což tedy vypovídá o tom, že ani v tomto kraji není vysoká dětská kriminalita.

Tab. 8: Asociační pravidla pro Středočeský kraj

ID	Středočeský kraj	podpora	spolehlivost
1	loupež spáchána mládeží od 7 do 12 hod.	50%	13%
2	loupež v neděli od 7 do 12 hod.	50%	13%
3	loupež v úterý od 7 do 12 hod.	42%	13%
4	loupež v pátek od 7 do 12 hod.	40%	13%

Podle vygenerovaných asociačních pravidel to vypadá, že ve Středočeském kraji děti loupeže páchají pouze v dopoledních hodinách. Procento případů je ale opět velmi nízké, takže i zde je dětská kriminalita nízká.

Tab. 9: Asociační pravidla pro kraj Hlavní město Praha

ID	Hlavní město Praha	podpora	spolehlivost
1	loupež spáchána mládeží v úterý	47%	18%
2	loupež spáchána mládeží od 7 do 12 hod.	46%	15%
3	loupež spáchána mládeží od 13 do 18 hod.	44%	18%
4	loupež spáchána mládeží od 19 do 24 hod.	44%	18%

Výsledky pro kraj Hlavní město Praha jsou ovšem trochu překvapivé. Vzhledem k počtu obyvatel a počtu městských částí, které byly do generování asociačních pravidel zahrnuty, by se daly očekávat vyšší hodnoty. Ale podle výsledků je Praha na stejné úrovni jako ostatní kraje, a oproti např. Ústeckému kraji je na tom, co se týče dětské kriminality, o dost lépe.

Tab. 10: Asociační pravidla pro kraj Vysočina

ID	Vysočina	podpora	spolehlivost
1	loupež spáchána mládeží v neděli	100%	33%
2	loupež spáchána mládeží v pondělí	100%	33%
3	loupež spáchána mládeží ve čtvrtek	66%	33%
4	loupež v pondělí od 19 do 24 hod.	66%	33%
5	loupež v sobotu od 13 do 18 hod.	66%	33%
6	loupež v neděli od 19 do 24 hod.	66%	33%

Při postupu na Moravu se kriminalita trochu zvýšila, je vidět, že ve 100% případech byla loupež v pondělí a neděli spáchána mládeží, z čehož ve třetině případů šlo o dětské pachatele. Je zde vidět, že činy byly spáchány v odpoledních a večerních hodinách.

Tab. 11: Asociační pravidla pro Olomoucký kraj

ID	Olomoucký kraj	podpora	spolehlivost
1	loupež v pondělí od 19 do 24 hod.	100%	45%
2	loupež spáchána mládeží od 19 do 24 hod.	100%	36%
3	loupež spáchána mládeží v neděli	100%	36%
4	loupež spáchána mládeží od 13 do 18 hod.	100%	36%
5	loupež v pondělí od 13 do 18 hod.	100%	36%
6	loupež spáchána mládeží v úterý	100%	36%

Zvyšující se „trend“ dětské kriminality s postupem na Moravu je potvrzen i v Olomouckém kraji, kde už se procento spáchaných loupeží pohybuje okolo 36% případů. Loupeže jsou páchany hlavně v odpoledních a večerních hodinách, jak je vidět v Tabulce 11.

Tab. 12: Asociační pravidla pro Moravskoslezský kraj

ID	Moravskoslezský kraj	podpora	spolehlivost
1	loupež v neděli od 7 do 12 hod.	100%	58%
2	loupež ve čtvrtek od 7 do 12 hod.	100%	62%
3	loupež spáchána mládeží ve čtvrtek	93%	62%
4	loupež v neděli od 19 do 24 hod.	93%	62%
5	loupež spáchána recidivistou v neděli	93%	62%
6	loupež spáchána mládeží v neděli	93%	62%
7	loupež ve čtvrtek od 19 do 24 hod.	93%	62%
8	loupež v neděli od 1 do 6 hod.	93%	58%
9	loupež v pondělí od 13 do 18 hod.	93%	58%
10	loupež v neděli od 13 do 18 hod.	92%	54%

Asi největší procento dětských loupeží je podle výsledků v Moravskoslezském kraji. Spolehlivost 62% už je číslo, vedoucí k zamyšlení nad tím, co by se dalo udělat pro snížení kriminality. Loupeže jsou páčány v rozmezí celého dne, podle nejvýše postavených asociačních pravidel, hlavně ve čtvrtek a v neděli.

Tab. 13: Asociační pravidla pro Jihomoravský kraj

ID	Jihomoravský kraj	podpora	spolehlivost
1	loupež ve středu od 13 do 18 hod.	100%	61%
2	loupež ve středu od 7 do 12 hod.	100%	61%
3	loupež ve středu od 19 do 24 hod.	100%	53%
4	loupež ve čtvrtek od 7 do 12 hod.	100%	53%
5	loupež ve čtvrtek od 13 do 18 hod.	100%	53%
6	loupež ve čtvrtek od 19 do 24 hod.	100%	53%
7	loupež v úterý od 13 do 18 hod.	100%	53%
8	loupež spáchána mládeží od 13 do 18 hod.	100%	61%
9	loupež spáchána mládeží ve středu	100%	61%
10	loupež spáchána mládeží v pondělí	100%	61%

Na jihu Moravy je vykazována také poměrně vysoká kriminalita, hlavně ve středu a ve čtvrtek v ranních a odpoledních hodinách. V 61% případů byl trestný čin spáchán dítětem, což, stejně jako v Moravskoslezském kraji, může vyvolávat otázky, co je v tomto kraji „špatně“ a jak tato čísla snížit.

Pro úplnost údajů ze všech krajů by zde měly být uvedeny ještě Pardubický a Zlínský kraj, avšak v těchto dvou krajích nebyla z poskytnutých dat zjištěna žádná dětská kriminalita. Pokud vezmeme v úvahu výsledky z celé Moravy, je Zlínský kraj opravdu velkou výjimkou.

Nastavení parametrů *Podpora* a *Spolehlivost* na tyto výsledky nemají žádný vliv, a to z toho důvodu, že v těchto krajích nebyla evidována Policií ČR žádná dětská kriminalita, proto nemohla být vygenerována žádná asociační pravidla. Aby v těchto krajích byly výsledky korektní, bylo by zapotřebí, aby bylo v těchto krajích více policejních útvarů, které by trestné činy evidovaly.

3.2 Loupeže spáchané dětmi v celé ČR

V následující tabulce jsou vypsána tři nejvýznamnější pravidla pro celou republiku. Je zde potvrzeno, že nejvíce činů páchají děti v odpoledních hodinách a více v týdnu než o víkendu. Těmto výsledkům se nejvíce přibližují hodnoty z Jihočeského, Karlovarského a Olomouckého kraje, zatímco nejvíc odlišné hodnoty, když nebudou vzaty do úvahy kraje, ve kterých nebyla zjištěna žádná kriminalita, jsou ve Středočeském kraji, Praze, Moravskoslezském a Ústeckém kraji.

Tab. 14: Asociační pravidla pro Českou republiku

ID	Česká republika	podpora	spolehlivost
1	loupež spáchaná mládeží v pondělí	64%	31%
2	loupež spáchaná mládeží od 13 do 18 hod.	63%	30%
3	loupež spáchaná mládeží v pátek	61%	30%

3.3 Loupeže spáchané mladistvými v krajích ČR

Aby bylo možné porovnat, jaký podíl v kriminalitě mládeže mají děti a jaký mladiství, jsou zde i pro ně uvedeny tabulky. V tabulkách je opět vypsáno maximálně 10 nejvýznamnějších pravidel.

Tab. 15: Asociační pravidla pro Jihočeský kraj

ID	Jihočeský kraj	podpora	spolehlivost
1	loupež spáchaná mládeží ve středu	100%	54%
2	loupež spáchaná mládeží od 19 do 24 hod.	100%	45%
3	loupež spáchaná mládeží od 1 do 6 hod.	100%	36%
4	loupež v pondělí od 19 do 24 hod.	100%	36%
5	loupež v sobotu od 19 do 24 hod.	100%	36%
6	loupež spáchaná mládeží v neděli	100%	36%

V této zájmové skupině už je vidět, že se procento případů trestné činnosti zvedlo a ve většině tabulek nebudou hodnoty spolehlivosti pod hranicí 30 %. V jihočeském kraji je vidět, že mladiství loupeže páchají převážně ve večerních a brzkých ranních hodinách a nejen v týdnu, ale i v sobotu a neděli.

Tab. 16: Asociační pravidla pro Plzeňský kraj

ID	Plzeňský kraj	podpora	spolehlivost
1	loupež spáchána mládeží v neděli	83%	41%
2	loupež spáchána mládeží od 1 do 6 hod.	83%	41%
3	loupež spáchána mládeží ve středu	83%	41%
4	loupež spáchána mládeží v úterý	80%	33%
5	loupež spáchána mládeží v sobotu	80%	33%

V Plzeňském kraji ani tentokrát není velká kriminalita, z tabulky je vidět, že nejvíc loupeží se stalo v úterý, ve středu a o víkendu. Je však na zvážení, jestli je 41% případů už významná hodnota, či nikoliv.

Tab. 17: Asociační pravidla pro Karlovarský kraj

ID	Karlovarský kraj	podpora	spolehlivost
1	loupež spáchána mládeží od 13 do 18 hod.	100%	80%
2	loupež spáchána mládeží v sobotu	100%	80%
3	loupež spáchána mládeží v pátek	100%	80%
4	loupež spáchána mládeží ve středu	100%	80%
5	loupež spáchána mládeží od 7 do 12 hod.	100%	80%
6	loupež v neděli od 13 do 18 hod.	100%	60%
7	loupež ve čtvrtek od 13 do 18 hod.	100%	60%
8	loupež v pátek od 19 do 24 hod.	100%	60%
9	loupež spáchána mládeží od 19 do 24 hod.	100%	60%
10	loupež spáchána recidivistou do 19 do 24 hod.	100%	60%

Kriminalita dětí v Karlovarském kraji byla téměř bezvýznamná, za to kriminalita mladistvých je jedna z nejvyšších v porovnání s ostatními kraji. Je vidět, že loupeže jsou páchány od ranních až do nočních hodin a pokud trestný čin spáchala mládež, jednalo se v 80% o mladistvého. Nejvíce loupeží se stalo od středy do soboty.

Tab. 18: Asociační pravidla pro Ústecký kraj

ID	Ústecký kraj	podpora	spolehlivost
1	loupež spáchána mládeží ve středu	100%	66%
2	loupež spáchána mládeží od 13 do 18 hod.	100%	66%
3	loupež spáchána mládeží od 19 do 24 hod.	100%	61%
4	loupež spáchána mládeží v pondělí	100%	61%
5	loupež spáchána mládeží v neděli	100%	61%
6	loupež v pondělí od 1 do 6 hod.	100%	61%
7	loupež ve středu od 1 do 6 hod.	100%	61%
8	loupež v neděli od 1 do 6 hod.	100%	55%
9	loupež v sobotu od 7 do 12 hod.	100%	55%

I v tomto kraji je vysoké procento spáchání loupeže mladistvými. V 66% případů spáchali trestný čin v odpoledních a večerních hodinách, v týdnu to pak bylo v 61% případů v brzkých ranních hodinách.

Tab. 19: Asociační pravidla pro Liberecký kraj

ID	Liberecký kraj	podpora	spolehlivost
1	loupež spáchána mládeží ve středu	100%	33%
2	loupež spáchána recidivistou ve středu	100%	33%
3	loupež ve středu od 13 do 18 hod.	100%	33%
4	loupež ve středu od 19 do 24 hod.	100%	33%
5	loupež ve čtvrtek od 13 do 18 hod.	100%	33%
6	loupež spáchána mládeží ve čtvrtek	100%	33%

V Libereckém kraji není procento loupeží mladistvých tak vysoké, z tabulky plyne, že se ve 33% případů staly loupeže ve středu a ve čtvrtek v odpoledních a večerních hodinách. Možná by stálo za to zvážit otázku, čím je to způsobeno, že zrovna v tyto dva dny je kriminalita vyšší.

Tab. 20: Asociační pravidla pro Královehradecký kraj

ID	Královehradecký kraj	podpora	spolehlivost
1	loupež ve středu od 13 do 18 hod.	100%	50%
2	loupež v pátek od 19 do 24 hod.	100%	50%
3	loupež spáchána mládeží v pátek	100%	50%
4	loupež v pátek od 13 do 18 hod.	100%	50%
5	loupež spáchána mládeží v úterý	100%	50%
6	loupež ve středu od 19 do 24 hod.	100%	50%

V Královéhradeckém kraji se trestná činnost mladistvých oproti dětem zvedla o 50%. Středa je opět dnem, který je nejrizikovější, v tomto kraji společně s pátkem. A jako ve většině ostatních krajů se nejvíc loupeží stalo v odpoledních a večerních hodinách.

Tab. 21: Asociační pravidla pro Pardubický kraj

ID	Pardubický kraj	podpora	spolehlivost
1	loupež spáchána mládeží od 19 do 24 hod.	100%	100%
2	loupež spáchána mládeží ve čtvrtek	100%	100%
3	loupež spáchána mládeží v pátek	100%	100%
4	loupež ve čtvrtek od 19 do 24 hod.	100%	60%
5	loupež ve středu od 7 do 12 hod.	100%	40%
6	loupež spáchána mládeží v pondělí	100%	40%

Dětští pachatelé v tomto kraji vyhledání nebyli, o to víc je zde pachatelů v řadách mladistvých. Ve 100% případů, u nichž byla jako pachatel zaznamenána mládež, šlo ve čtvrtek, v pátek a ve večerních hodinách o mladistvého.

Tab. 22: Asociační pravidla pro Středočeský kraj

ID	Středočeský kraj	podpora	spolehlivost
1	loupež spáchána mládeží od 19 do 24 hod.	100%	39%
2	loupež spáchána mládeží od 1 do 6 hod.	100%	39%
3	loupež spáchána mládeží v pátek	100%	34%
4	loupež spáchána mládeží v úterý	100%	39%
5	loupež spáchána mládeží od 13 do 18 hod.	100%	34%

Ve Středočeském kraji není, jak je vidět, kriminalita mládeže, tedy společně dětí i mladistvých vysoká. Ani v této tabulce nejsou k nalezení vysoké hodnoty. Z tabulky vyplývá, že loupeže jsou páhány hlavně ve večerních a nočních hodinách, hlavně v úterý a v pátek.

Tab. 23: Asociační pravidla pro kraj Hlavní město Praha

ID	Hlavní město Praha	podpora	spolehlivost
1	loupež spáchána mládeží ve středu	86%	29%
2	loupež spáchána recidivistou ve čtvrtek	86%	29%
3	loupež spáchána mládeží v neděli	86%	29%
4	loupež spáchána mládeží od 1 do 6 hod.	82%	31%
5	loupež spáchána mládeží v pondělí	77%	31%

V případě Prahy jsou opět získány překvapivé hodnoty, maximálně ve 31% případů spáchaných mládeží šlo o trestný čin mladistvého. Nejvíce loupeží bylo spácháno v pondělí, ve středu a ve čtvrtek, ve 31% případů od 1 do 6 hodiny ranní.

Tab. 24: Asociační pravidla pro kraj Vysočina

ID	Vysočina	podpora	spolehlivost
1	loupež spáchána recidivistou v pátek	100%	33%
2	loupež spáchána recidivistou v sobotu	100%	33%
3	loupež v sobotu od 1 do 6 hod.	100%	33%
4	loupež v pátek od 1 do 6 hod.	100%	33%
5	loupež spáchána mládeží od 1 do 6 hod.	100%	33%

Z výše uvedené tabulky je zjevné, že na Vysočině se nejvíce loupeží stává v pátek a v sobotu v nočních hodinách. Tato čísla by mohly ovlivňovat diskotéky a jiné zábavy mladistvých, které se v tuto dobu konají.

Tab. 25: Asociační pravidla pro Olomoucký kraj

ID	Olomoucký kraj	podpora	spolehlivost
1	loupež spáchána mládeží v sobotu	83%	45%
2	loupež spáchána mládeží v pondělí	83%	45%
3	loupež spáchána mládeží ve čtvrtek	83%	45%
4	loupež spáchána mládeží v pátek	80%	36%

Jak je vidět, v Olomouckém kraji 45% případů loupeží spáchali mladiství, nejčastěji v pondělí, ve čtvrtek a v sobotu.

Tab. 26: Asociační pravidla pro Moravskoslezský kraj

ID	Moravskoslezský kraj	podpora	spolehlivost
1	loupež v sobotu od 13 do 18 hod.	87%	58%
2	loupež v úterý od 13 do 18 hod.	87%	58%
3	loupež v pondělí od 13 do 18 hod.	86%	54%
4	loupež v neděli od 13 do 18 hod.	85%	50%
5	loupež v pondělí od 7 do 12 hod.	84%	45%
6	loupež spáchána mládeží od 7 do 12 hod.	84%	45%
7	loupež v pátek od 7 do 12 hod.	84%	45%
8	loupež ve středu od 7 do 12 hod.	84%	45%
9	loupež spáchána recidivistou od 7 do 12 hod.	84%	45%

V Moravskoslezském kraji je podle všeho méně trestných činů spáchaných mladistvými než dětmi. Procento případů se o trochu snížilo a je zde vidět i překvapivé rozmezí hodin spáchání loupeže, ve většině případů v dopoledních a odpoledních hodinách, během celého týdne.

Tab. 27: Asociační pravidla pro Zlínský kraj

ID	Zlínský kraj	podpora	spolehlivost
1	loupež spáchána mládeží od 1 do 6 hod.	100%	28%
2	loupež spáchána mládeží v neděli	100%	28%
3	loupež spáchána mládeží ve středu	100%	28%

Ve Zlínském kraji nebyli v řadách dětí evidováni žádní pachatelé, a také v řadách mladistvých je počet trestných činů celkem nízký. Loupeže se staly pouze ve středu a v neděli a jejich procentuální vyjádření je nižší než 30 procent ze všech případů.

Tab. 28: Asociační pravidla pro Jihomoravský kraj

ID	Jihomoravský kraj	podpora	spolehlivost
1	loupež spáchána mládeží v pondělí	100%	61%
2	loupež spáchána mládeží ve středu	100%	61%
3	loupež spáchána mládeží od 13 do 18 hod.	100%	61%
4	loupež ve středu od 7 do 12 hod.	100%	61%
5	loupež spáchána mládeží v pátek	100%	61%
6	loupež ve středu od 13 do 18 hod.	100%	61%
7	loupež v úterý od 13 do 18 hod.	100%	53%
8	loupež ve středu od 19 do 24 hod.	100%	53%
9	loupež ve čtvrtek od 7 do 12 hod.	100%	53%
10	loupež ve čtvrtek od 19 do 24 hod.	100%	53%

Nejvyšší procento účasti mladistvých na loupeži, co se týče Moravy, je určité v Jihomoravském kraji. Přes den jsou evidovány přestupky téměř u 61% mladistvých z řad mládeže, ve středu a ve čtvrtek jsou zaznamenány loupeže také ve večerních hodinách, kde se na nich podíleli mladiství z 53%.

3.4 Loupeže spáchané mladistvými v celé ČR

Pravidla vygenerovaná pro mladistvé se v porovnání s pravidly pro děti moc nezměnila. Je pravda, že hranice počtu případů se posunula o trochu výše, ale vzhledem k tomu, že jsou tyto hodnoty průměrem ze všech obcí, je hranice 40% případů spáchaných mladistvými poměrně přijatelná.

Tab. 29: Asociační pravidla pro Českou republiku

ID	Česká republika	podpora	spolehlivost
1	loupež spáchaná mládeží ve středu	90%	43%
2	loupež spáchaná mládeží od 1 do 6 hod.	89%	40%
3	loupež spáchaná mládeží od 13 do 18 hod.	86%	41%
4	loupež spáchaná mládeží v sobotu	86%	40%
5	loupež spáchaná mládeží v pondělí	86%	42%
6	loupež spáchaná mládeží v úterý	86%	40%
7	loupež spáchaná mládeží od 19 do 24 hod.	86%	41%

Z těchto dat tedy vyplývá, že nejvíce činů spáchaných mladistvými je spáchano v pondělí, úterý a ve středu. Hodnoty u některých krajů jsou oproti hodnotám pro celou Českou republiku nadprůměrné, přičemž zde může být jako příklad uveden Karlovarský, Ústecký, Jihomoravský a Moravskoslezský kraj. V Libereckém, Středočeském a Zlínském kraji se i v případě mladistvých drží kriminalita mladistvých na nízkých hodnotách.

4 Závěr

Na konkrétní data poskytnutá Policií ČR byla aplikována metoda asociace a díky ní bylo získáno mnoho informací o vztazích mezi daty, někdy pro mě velmi překvapivých. Např. kriminalita mládeže v Praze vůbec není tak vysoká, jak je všeobecně předpokládáno, nebo Zlínský kraj je, co se týče vytyčeného problému, nejbezpečnější na celé Moravě.

Na začátku aplikace asociace na reálná data bylo položeno několik otázek, na něž jsem se snažila najít odpověď. Zde je místo pro tyto odpovědi.

1. Souvisí kriminalita dětí a mladistvých s krajem, ve kterém žijí?

Souvisí, bylo dokázáno, že procento případů kriminality mládeže je v jednotlivých krajích velmi odlišné. Jak se tyto hodnoty liší, je možné zjistit v kapitole 6.

2. Jak se liší asociační pravidla vygenerovaná pro celou Českou republiku od pravidel pro jednotlivé kraje?

Počty případů loupeží, získané pro celou republiku jsou poměrně nízké. V některých krajích jsou výsledky odlišné od průměrných hodnot. Například v Ústeckém kraji jsou loupeže dětí evidovány v 50% případů, zatímco průměrné hodnoty jsou 31% případů.

3. Je nějaká souvislost mezi dnem v týdnu a spácháním loupeže dětmi a mladistvými?

Z asociačních pravidel bylo patrné, že děti páchají trestné činy hlavně v týdnu. Mladiství páchají loupeže hlavně v sobotu, v týdnu většinou jen ve středu a ve čtvrtek.

4. Spáchají děti loupež spíše ve dne nebo v noci?

Jak bylo patrné, ve většině krajů spáchaly děti trestný čin během dne.

5. Pokud spáchá trestný čin loupeže mládež, jaký podíl v celkovém součtu mají děti a mladiství?

Asi není překvapivé, že na tuto otázku odpovídám, že hlavní podíl v celkovém součtu mají mladiství. Důvodem může být jak puberta, tak alkohol na diskotékách či potřeba většího finančního obnosu.

Literatura

- [1] *Oracle Spatial Developer's Guide 11g Release 1.*, Oracle corp. Dostupné [online]:
http://download.oracle.com/docs/cd/B28359_01/appdev.111/b28400.pdf
- [2] *Oracle Data Mining Concepts 11g Release 1.* Oracle corp. Dostupné [online]:
http://download.oracle.com/docs/cd/B28359_01/datamine.111/b28129.pdf
- [3] Shekhar,S.,Zhang,P. a col.: *Trends in Spatial Data Mining*. Dostupné [online]:
http://www.spatial.cs.umn.edu/paper_ps/dmchap.pdf
- [4] Nemrava, M., Pospíšil J.: *Dolování dat a jeho aplikace*. Dostupné [online]:
<http://axpsu.fpf.slu.cz/~sos10um/trendy/DM.pdf>
- [5] *Územně identifikační registr ÚIR-ZJS, Český statistický úřad*. Dostupné [online]:
http://www.liberec.czso.cz/csu/rso.nsf/i/prohlizec_uir_zsj
- [6] Witte,E.: *Spatial Data Mining*. Dostupné [online]:
<http://www1.in.tum.de/teaching/ws01/CBP-Hauptseminar/Presentations/SpatialDataMining-Pres.pdf>
- [7] Haberstroh, R.: *ODMr 11g Tutorial for OTN*. Dostupné [online]:
<http://www.oracle.com/technology/products/bi/odm/odminer.html>
- [8] *Wikipedie otevřená encyklopedie*. Dostupné [online]: <http://www.wikipedia.org>
- [9] *PostgreSQL*. Dostupné [online]: <http://www.pgsql.cz/index.php/PostgreSQL>
- [10] *RapidMiner software Informer*. Dostupné [online]:
<http://rapidminer.software.informer.com/>
- [11] *MySQL*. Dostupné [online]: <http://www.mysql.com/>

- [12] *Weka, Data Mining with Open Source*. Dostupné [online]:
<http://www.cs.waikato.ac.nz/ml/weka>
- [13] Danel, R.: *Dolování dat*. Dostupné [online]:
<http://homel.vsb.cz/~dan11/isis/Danel%20-%20IS%20-%20Dolovani%20dat.pdf>
- [14] *Geomarketing*. http://geomarketing.eu/cz/xc4_geom.htm
- [15] *Microsoft SQL Server*. Dostupné [online]:
<http://www.microsoft.com/cze/windowsserversystem/sql/default.mspx>
- [16] ARCDATA PRAHA: ArcČR 500. Dostupné [online]:
<http://old.arccdata.cz/data/arccr>
- [17] Janečka, K.: *Modelování konzistentní báze geodat na úrovni datového modelu katastru nemovitostí*. Plzeň, 2009.
- [18] *Oracle Map Builder*. Dostupné [online]:
<http://www.oracle.com/technology/software/products/mapviewer/index.html>