



TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky
a mezioborových studií ■

Prevence kriminality s podporou dataminingových modelů

Diplomová práce

M13000203

Studijní program: N2612 – Elektrotechnika a informatika

Studijní obor: 1802T007 – Informační technologie

Autor práce: **Bc. Jan Šrámek**

Vedoucí práce: RNDr. Klára Císařová, Ph.D.





TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Mechatronics, Informatics
and Interdisciplinary Studies ■

Crime Prevention supported by Datamining models

Diploma thesis

M13000203

Study programme: N2612 – Electrical Engineering and Informatics

Study branch: 1802T007 – Information technology

Author: **Bc. Jan Šrámek**

Supervisor: RNDr. Klára Císařová, Ph.D.



ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Jan Šrámek**
Osobní číslo: **M13000203**
Studijní program: **N2612 Elektrotechnika a informatika**
Studijní obor: **Informační technologie**
Název tématu: **Prevence kriminality s podporou dataminingových modelů**
Zadávající katedra: **Ústav mechatroniky a technické informatiky**

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

1. Seznamte se s aktuálním stavem podpory prevence kriminality v ČR.
2. Vytipujte vhodné veřejně dostupné datové zdroje a data, která by vhodně doplnila standardně sbíraná data KPČR.
3. Analyzujte typy kriminálního a podvodného jednání, kdy může data mining přispět k prevenci.
4. Postavte model pro predikci prevence nad vybranými daty, spolupracujte s analytiky PČR.

Rozsah grafických prací: dle potřeby dokumentace

Rozsah pracovní zprávy: cca 40–50 stran

Forma zpracování diplomové práce: tištěná/elektronická

Seznam odborné literatury:


- [1] Olivia Parr Rud: Datamining, Computer Press, a.s., 2006
- [2] Yong Yin, Ikou Kaku, Jiafu Tang: Data Mining, Springer London Ltd, 2011
- [3] Hendl J.: Přehled statistických metod zpracování dat, Portál, s.r.o. 2006
- [4] <http://www.msps.cz/data-mining/>

Vedoucí diplomové práce: **RNDr. Klára Císařová, Ph.D.**
Ústav mechatroniky a technické informatiky

Datum zadání diplomové práce: **10. října 2015**
Termín odevzdání diplomové práce: **16. května 2016**


prof. Ing. Václav Kopecký, CSc.
děkan




doc. Ing. Milan Kolář, CSc.
vedoucí ústavu

V Liberci dne 10. října 2015

Prohlášení

Byl jsem seznámen s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé diplomové práce pro vnitřní potřebu TUL.

Užiji-li diplomovou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Diplomovou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím mé diplomové práce a konzultantem.

Současně čestně prohlašuji, že tištěná verze práce se shoduje s elektronickou verzí, vloženou do IS STAG.

Datum: 16.5.2016

Podpis: 

Poděkování

Rád bych poděkoval vedoucí práce RNDr. Kláře Císařové, Ph.D. za její ochotu při pomoci s prací, doporučení materiálů a konstruktivní připomínky. Dále bych rád poděkoval své přítelkyni, která mě vždy velmi podporovala. Stejně tak jako mým rodičům, kteří mi celé studium umožnili.

Abstrakt

Tato práce se zabývá využitím dataminingu pro predikci kriminality. Jejím cílem je navrhnout modely, které by pro vybrané typy trestných činů tento problém řešily. Součástí práce je rešerše možných datových zdrojů, které by mohly při analýze vhodně doplnit data dnes běžně využívaná analytickými odděleními Policie České republiky.

Dále byl navržen a implementován nástroj, který by umožnil policejním složkám sbírat data přímo z terénu pomocí mobilního zařízení a ukládat je do vytvořené databáze. Byla také vytvořena webová aplikace sloužící jako dispečink a administrace. S použitím analytického open-source nástroje Knime byly navrženy ukázkové modely pro možnou predikci kriminality.

Klíčová slova

Prevence kriminality, dataming, shluková analýza

Abstract

This thesis focuses on the use of data mining to predict crime. The aim of this thesis is to design models that would solve this issue for selected types of crime. The thesis also contains research of possible data sources that could aptly complement the data that are nowadays commonly used by analytical departments of the Police of the Czech Republic.

Furthermore, a tool that was also designed and implemented would enable police forces to collect data directly from the field using mobile devices and store them in a database. Next, a web application was developed within the scope of this thesis. This application can serve for dispatch and administrative purposes. Last, sample models for crime prediction were designed using an open-source analytical tool Knime.

Keywords

Crime prediction, dataming, cluster analysis

Obsah

Úvod	11
2 Prevence kriminality v ČR	12
2.1 Typy prevence	12
2.1.1 Primární prevence	13
2.1.2 Sekundární prevence	13
2.1.3 Terciární prevence	14
2.1.4 Sociální prevence	14
2.1.5 Situační prevence	14
2.2 Souhrné informace	15
2.3 Projektová podpora prevence kriminality	16
2.3.1 Projekt Mapy budoucnosti	16
2.3.2 Bezpečný Kolín	17
3 Data pro prediktivní modely prevence kriminality	20
3.1 Volně dostupné datové zdroje	20
3.2 Otevřená data	23
3.3 Další data, která mohou přispět k predikci	23
3.4 Formáty dat	24
4 Typy kriminálního jednání a datamining	25
4.1 Trestní kriminalita a datamining	25
4.2 Ekonomická kriminalita a datamining	26
4.3 Internet a trestní jednání	26
5 Vytvořený systém pro predikci výskytu trestné činnosti	27
5.1 Teoretický podklad	27

5.1.1	Shluková analýza	27
5.1.2	Koeficienty asociace objektů	30
5.1.3	Metriky	30
5.1.4	Algoritmus k-means	31
5.1.5	Popis systému	32
5.2	Mobilní aplikace	32
5.3	Databáze	34
5.4	Webová aplikace	35
5.5	Knime	38
5.6	Bezpečnost	39
5.7	Použitá data	39
5.8	Úloha Vloupání	42
5.9	Úloha kapesní krádeže	47
5.10	Úloha krádeže vozidel	49
	Závěr	54

Seznam obrázků

2.1	Trestná činnost v Kolíně znázorněná pomocí metody hot-spots	18
5.1	Grafické znázornění hierarchického shlukování	28
5.2	Grafické znázornění nehierarchického shlukování	28
5.3	Diagram vytvořeného systému	32
5.4	Interface mobilní aplikace	34
5.5	Schéma navržené databáze	35
5.6	Ukázka přehledu hlídek v terénu	37
5.7	Všechny trestné činy	42
5.8	Rozdělení případů podle časových úseků	43
5.9	Ráno	44
5.10	Den	44
5.11	Večer a noc	44
5.12	Ukázka proudu pro analýzu vloupání a nasazení hlídek	45
5.13	Výsledek analýzy souvisejících vloupání	46
5.14	Histogram znázorňující množství kapesních krádeží podle týdnů v roce	47
5.15	Zobrazení krádeží rozdělených dle kritických období	48
5.16	Histogram krádeží obarvený podle oblastí	49
5.17	Zobrazení rizikových oblastí krádeží automobilů	50
5.18	Zastoupení počasí při krádeži	51
5.19	Histogram četnosti krádeží vozu dle roku výroby	52

Úvod

Boj proti zločinu je stejně starý jako kriminalita sama. Proto v dnešní době informatiky narůstá snaha využít dostupné technologie také pro tyto účely. Díky technickému pokroku je dnes možné získávat data z terénu v reálném čase a v tento okamžik data také analyzovat. Při vhodném způsobu shromáždění informací slouží datamining k odhalování na první pohled skrytých spojitostí. Je tak možné zjistit souvislosti mezi jednotlivými trestnými činy, či okolnostmi události. Po celém světě vznikají projekty zabývající se zefektivněním práce policie pomocí moderních matematických analytických nástrojů. Právě zahraniční zkušenosti ukazují, že na základě analýz a predikcí lze snížit obecnou kriminalitu až o 30%. Konkrétně tento závěr pochází ze statistik policejních orgánů města Memphis ve Spojených státech. Zde se místní policie rozhodla provést zásadní změny ve způsobu práce a vytvořili program Blue CRUSH. Tento zahraniční program jehož myšlenkou je zefektivnění práce policejních složek s využitím dataminingu byl hlavní inspirací pro praktickou část této diplomové práce.

2 Prevence kriminality v ČR

2.1 Typy prevence

Prevence kriminality je významná část bezpečnostní politiky ČR. Preventivní politika řeší bezpečnost společnosti v budoucnosti. Jde o to, kriminálním činům bránit účinnými opatřeními. Předvídat páchaní trestních činů a jiného nežádoucího jednání v místě a čase a zasáhnout nerepresivními prostředky tak, aby se pravděpodobnost páchaní trestních činů snížila co nejvíce. Naproti tomu trestní politika řeší trestní činy, které se již staly. Jejich vyřešení a nasazení represe má odradit od páchaní trestních činů a tedy v tomto smyslu má také preventivní funkci. Zajímavým rysem je pohled na subjekty, které se na trestní a preventivní politice podílejí. Trestní politiku vykonávají orgány činné v trestním řízení, tj. policie, státní zastupitelství, soudy a vězeňství. Jejich role jsou právně přesně vymezené. Preventivní politiku realizuje mnohem více subjektů. V prevenci působí už vyjmenované orgány a dále orgány státní správy, sdružení občanů, církve, analytická oddělení orgánů státní správy i občané svou angažovanou účastí na životě občanské společnosti. Kriminální činnost je velmi široký okruh jednání a některé speciální druhy trestní činnosti mohou řešit pouze týmy expertů, kriminalistů a kontrolních orgánů ve spolupráci s profesionály. Za zmínku například stojí problém organizovaného zločinu, terorizmu, nebo hospodářské kriminality, počítačového rizikového chování, kyberšikany apod.

Prevence kriminality, z jiného úhlu pohledu, je částí vědy zvané kriminologie. Kriminologie je věda o kriminalitě, a její kontrole. Je multidisciplinární a do jejího obsahu se promítají poznatky z jiných vědních oborů. Jen pro ilustraci jsou to například poznatky psychologie, psychiatrie, medicíny i sociologie pro zkoumání a vysvětlování chování pachatelů a také obětí. Hledání souvislostí a trendů kriminálních chování je spojeno se statistikou a dnes i s algoritmy dataminingu a umělé inteligence. Nelze vynechat souvislost s právním řádem. Výsledky kriminalistického zkoumání by měly sloužit zákonodárcům při formulování a schvalování zákonných úprav. Například

současně velmi diskutovaná trestní odpovědnost mladistvých je již dlouhou dobu řešena na půdě parlamentu ČR. Bohužel odborníky garantované názory jsou prezentované veřejnosti mnohém méně než politicky motivované často laické názory. cit[cis1]

2.1.1 Primární prevence

Pracuje s celou populací, s dospělými i dětmi. Jde o aktivity výchovné, vzdělávací, různé poradenské služby pro nejširší veřejnost. Společnost musí dostávat relevantní odborné informace, musí mít znalost základních právních povinností, vědomí trestné odpovědnosti atd. Velký důraz je kladen na pozitivní ovlivňování hodnotových měřítek především u dětí a mládeže. Řešení je v uvážené nabídce využívání volného času a to jak nabídkou kulturních příležitostí, tak sportovních a jiných aktivit. Primární prevenci fakticky dělá rodina, škola a veřejnoprávní media a také orgány veřejné správy. Neformálně se na ní podílejí také různá zájmová sdružení a spolky.

2.1.2 Sekundární prevence

Jedná se o prevenci pracující s rizikovými skupinami, u kterých je pravděpodobnost páchání trestné činnosti nebo skupinami, které mohou být potencionálními oběťmi. Práce je selektivní podle věku, hrozícího nebezpečí, sociálně patologických jevů, prostředí a dalších kriminogenních situací. Na vysvětlení uvedu některé skupiny, kterým musí být věnovaná specializovaná sociální i vzdělávací péče. Jedná se například o lidi, kteří jsou drogově závislí, alkoholově závislí, gambleři, sprejeři, agresoři – výtržníci, vandalové, rasisti, extrémisti a další. Sekundární prevence je prováděná na úrovni státu, krajů, regionů i místních orgánů příslušnými pracovišti. Pro ilustraci:

Republikový výbor pro prevenci kriminality je orgán ministerstva vnitra, kraje mají manažery prevence kriminality, podobně mají své manažery i města. Aktuální seznamy jsou přílohy 2,3,4 této práce. Další funkce plní ministerstvo práce a sociálních věcí, ministerstvo spravedlnosti i ministerstvo zdravotnictví. Mnoho funkcí prevence plní i odbory sociálních věcí dalších státních orgánů na všech úrovních státní správy. Dále fungují různá krizová centra například pro drogově či jinak závislé, mnoho občanských sdružení i výzkumných organizací, které studují příčiny kriminogenních situací a pomáhají hledat postupy prevence.

2.1.3 Terciární prevence

Se zaměřuje na pachatele, oběti a na lokality, které již byly kriminalitou zasaženy. Lokality je potřeba kriminality zbavit, nebo ji alespoň limitovat. Pachatele je potřeba resocializovat, bránit recidivě trestné činnosti a hledat cesty k návratu kriminálně poznamenaných osob do života. Hledat možnosti pracovních uplatnění pachatelů po skončení výkonu trestu, nabídnout jim na trhu uplatnitelné rekvalifikace, poskytnout obětem i pachatelům sociální a rodinné poradenství a mnoho dalších potřebných funkcí souvisejících s životem odsouzených. Odpovědné orgány i v této oblasti prevence jsou výše uvedená ministerstva a orgány státní správy. Vzdělávání pachatelů ve věznicích spadá pod ministerstvo školství. Na prevenci kriminality je někdy vhodné dívat se pohledem sociální a situační prevence.

2.1.4 Sociální prevence

Tento typ prevence se zabývá zabráněním sociálnímu vyloučení osob ze společnosti. Z hlediska obětí jde o pomoc výchovou a poradenstvím k bezpečnému chování a předcházení kritickým situacím a to s ohledem na různý typ kritických situací. Ohrožená osoba, která projde psychologickým a právním poradenstvím, případně tréninkem v obranných strategiích, je připravená se účinně bránit, případně se kriminálními situacím vyhnout. Cílem tedy je ochrana společnosti před vznikem a šířením společensky nepřijatelných jevů.

2.1.5 Situační prevence

Vychází ze znalostí výskytu určité kriminality v místě a čase. Z historie a zdokumentované trestné činnosti odvozuje pravděpodobnost výskytu jistého druhu kriminálních činů. Následně organizačními či technickými opatřeními se snaží bránit výskytu kriminality. V praxi to znamená například instalaci kamerových systémů nebo nasazení policejních hlídek podle výsledků analýz. Patří sem také různé technické bezpečnostní prvky pro zabezpečení majetku, nebo bezpečnostní agentury, ideálně ve spojení s policií ČR.

2.2 Souhrné informace

Na prevenci kriminality se podílí vláda vytvářením preventivní politiky s poradním orgánem, který se jmenuje Republikový výbor pro prevenci kriminality. Ve výboru jsou zastoupené relevantní resorty. Pro ilustraci odborného meziresortního zastoupení je složení výboru následující:

- ředitel odboru bezpečnostní politiky a prevence kriminality, Ministerstvo vnitra ČR,
- odbor vzdělávání, sekce personální, Ministerstvo obrany ČR,
- první náměstek policejního prezidenta, Policejní prezidium ČR,
- odbor sociálních služeb a sociálního začleňování, Ministerstvo práce a sociálních věcí,
- odbor ochrany práv a dětí a náhradní rodinné péče, Ministerstvo práce a sociálních věcí,
- legislativní odbor, Ministerstvo spravedlnosti,
- odbor vzdělávání, oddělení prevence a speciálního vzdělávání Ministerstvo školství, mládeže a tělovýchovy,
- odbor financování kapitol státního rozpočtu II, Ministerstvo financí ČR,
- Nejvyšší státní zastupitelství,
- zástupce ředitele odboru výkonu vazby a trestu, Generální ředitelství vězeňské služby,
- ředitel Institutu pro kriminologii a sociální prevenci,
- oddělení kanceláře Rady vlády pro záležitosti romské menšiny a sekretariátu Rady vlády pro národnostní menšiny, Úřad vlády ČR,
- Národní monitorovací středisko pro drogy a závislosti Úřad vlády ČR,
- Ředitelka, Probační a mediační služba,
- Okresní soud Rychnov nad Kněžnou,

- ředitel odboru trestní politiky, vězeňství, probace a mediace, Ministerstvo spravedlnosti,

Na nižších úrovních spolupracují na prevenci příslušné odbory veřejné správy, policie, nevládní organizace a další instituce jako spolky, občanská sdružení apod. Na krajské úrovni se vytváří programy prevence kriminality. Za jejich realizaci nesou odpovědnost obecní zastupitelstva, odbor prevence kriminality. Ministerstvo vnitra pak poskytuje metodickou a konzultační pomoc a dotační prostředky na realizaci preventivních projektů. Cílem preventivní politiky je bezpečné prostředí a zvyšující se pocit bezpečí občanů. Přispět ke snížení míry a závažnosti trestné činnosti lze mnoha způsoby. Příspěvkem k situační prevenci je i tato práce.

2.3 Projektová podpora prevence kriminality

Prevence kriminality ve svých výstupech představuje soubor mnoha různých opatření zvyšujících bezpečnost společnosti jak bylo popsáno výše. Významně je prevence podpořena realizací řady projektů, které mají výzkumný, vzdělávací i aplikační charakter. Některé z nich byly inspirací pro tuto práci.

2.3.1 Projekt Mapy budoucnosti

Projekt byl financován z prostředků Evropského sociálního fondu v Operačním programu Lidské zdroje a zaměstnanost a ze státního rozpočtu ČR. Na projektu spolupracoval odbor prevence kriminality ministerstva vnitra a výzkumná organizace ACCENDA – Centrum pro vědu a výzkum od roku 2014. Šlo o to, hledat a poznávat moderní nástroje zpracování dat na analýzu a predikci kriminality s cílem zvýšit efektivitu a kvalitu výkonu veřejné správy v oblasti prevence kriminality. Výstupy tohoto projektu byly prezentované na dvoudenní mezinárodní konferenci „Mapy budoucnosti,“ které jsme se mohli ve 3-členném týmu FM zúčastnit. Tato konference byla cenným zdrojem informací nejen o současných trendech zpracování sofistikovaných analýz příčin, souvislostí, vazeb a predikce očekávaného vývoje kriminality v ČR, ale byla představena celá řada evropských a zámořských řešení prevence kriminality. Potvrdila, že ve vyspělých demokratických společnostech na analýzách úzce spolupracuje akademická sféra a bezpečnostní policejní složky. Dalším zjištěním byl fakt, že projekt Mapy budoucnosti je v oblasti predikčních analýz v ČR průkopníkem a zahraniční znalosti a zkušenosti představují především motivaci, inspiraci, či také

zdroj nových postupů pro řešení kriminality. Závěry a doporučení vycházejí zejména ze studia nasazení predikčních analýz v zemích Evropské unie a Spojených států amerických. Dnes již existují dílčí řešení pro vytváření map kriminality ve tvaru zobrazení „hot spot“ a „cold spot“ míst. Jsou to místa s nejvyšší a nejnižší intenzitou výskytu vybraného kriminálního jevu. Ze zkoumání časových závislostí, analýza napoví, která místa jsou riziková, s jakou pravděpodobností se tam bude vyskytovat sledovaný kriminální jev a kdy se takový čin dá očekávat. Analýzy mohou také pomoci rozšířovat další otázky, které si kriminalisti pokládají. Například: Proč právě tato lokalita, proč právě tento čas, proč tímto způsobem, proč tato oběť a tento pachatel. Souvislosti mohou být ukryté v datech, které má policie k dispozici. K těmto možnostem předcházení trestné činnosti obecně panuje jistá nedůvěra. Na stránkách projektu Mapy budoucnosti probíhá anketa mapující využívání opatření prevence kriminality, která jsou v současné době k dispozici. Z aktuálního výsledku vyplývá, že 41% respondentů se hlásí k využívání podobných opatření prevence, 29% pravděpodobně nemá důvěru v současná opatření a nevyužívá je a 30% by chtělo s prevencí pracovat, ale neví jak. Celkem hlasovalo 2716 lidí. Tato čísla mají jen ilustrativní význam, protože hlasování není autorizované pro relevantní osoby, řešitelé projektu, případně kriminalisty a analytiku. V neformálních diskusích se nám potvrdilo, že řada přítomných skutečně o nasazení pochybuje. Pochybnosti mají kořeny ve sníženém personálním stavu PČR a příliš přetížených policistech. Nová technologie potřebuje čas na vývoj, implementaci a také nasazení a pilotní provoz. Mapy budoucnosti měly posunout znalost a ukázat, že i v ČR existují první lokální řešení, která využívají nové postupy a metody.

Na konferenci bylo představeno unikátní řešení prevence kriminality v Kolíně. Ve Středočeském kraji se podařilo spojit relevantní instituce, které na prevenci mají zásadní zájem – Policie ČR, Městská policie a město Kolín. Projekt má název Bezpečný Kolín.

2.3.2 Bezpečný Kolín

Tento zajímavý projekt chce výrazným způsobem eliminovat pouliční kriminalitu, hlavně krádeže vloupáním a krádeže ostatní, ale řeší i kriminalitu a její prevenci jako celek.

Přístup k řešení kriminality nově deklaruje spolupraci místních úřadů a samosprávy. Zavadové osoby jsou sledovány a to zejména tak, že se pravidelně kontrolují vytipovaná kritická místa – bytovny, sběrný, nádraží vlakové i autobusové. Výkon

služby pracovních hlídek vychází z určení „hot spot“, míst, kde je evidovaný výskyt kriminálního chování. Zahraniční zkušenosti [podle závěrečné zprávy projektu Mapy budoucnosti] potvrzují, že pachatelé se cca v 25 % vrací do těchto míst, kde byla v minulosti trestná činnost páčána nebo tato místa trestnou činností „přitahují“ svými charakteristikami. Mohou to být bary, hospody, herny, nádraží, zastávky autobusů, sběrný v noci parky apod. Geograficky analytický systém dovoluje data různě filtrovat a podávat aktuálně potřebný výsledek analýz. Mapové podklady slouží potom pro řízení zásahů a dohledu. Systém je propojený s občany tak, aby byl sběr dat, podnětů a informací k jakékoliv nezákonné činnosti co nejlepší.

Podle článku prezentovaném na konferenci [5] celkový pokles trestné činnosti v městě Kolín byl oproti roku 2014 o 39 %, oproti roku 2013 je pokles o 50 %, u majetkové trestné činnosti činí pokles oproti roku 2014 o 57 %, oproti roku 2013 je to pokles o 68 %, u krádeží vloupáním byl registrován pokles oproti roku 2014 o 56 %, oproti roku 2013 se jedná o pokles o 85 % u krádeží prostých činí pokles oproti roku 2014 o 64 %, oproti roku 2013 je pokles o 71%. To jsou velmi nadějná čísla a potvrzují, že investice do projektu má i ekonomický efekt.



Obrázek 2.1: Trestná činnost v Kolíně znázorněná pomocí metody hot-spots
Zdroj: Policie ČR, Městská policie Kolín

V prezentaci bylo konstatováno, že je potřeba zlepšit sběr dat při šetření na místě trestných činů, při kontrolách osob v terénu a to tak, aby policista měl k dispozici tablet s aplikací, která mu umožní data vkládat do systému v terénu a také data v terénu získávat. Aplikace by měla poskytovat i pracovní řídicí informace. Současný stav podle dokumentu [5] je takový, že policista vkládá data do systému až v kan-

celáři z kontrolního papírového záznamu. Očekávané přínosy budou v efektivnějším využívání času, okamžitě po zjištění odeslaná informace, ať už o osobě, věci, místě, ovlivní rychlost šetření a zvýší se i počet policistů v ulicích.

3 Data pro prediktivní modely prevence kriminality

3.1 Volně dostupné datové zdroje

Tato část diplomové práce je zaměřena na průzkum dostupných resp. existujících relevantních dat. Jedná se o data, která by mohla doplnit běžně sbíraná data Policií ČR a přidat tím do analýz a predikčních modelů novou informaci. Byly hledány datové zdroje institucí, které mohou mít vhodné informace, veřejné dostupné rejstříky, ministerské agendy a další dostupné informace, které by mohly zlepšit kvalitu predikčních modelů.

Data Úřadu práce

Na <https://portal.mpsv.cz/sz/download> lze dohledat tyto data:

- volná místa Úřadu práce ČR v HTML nebo XML formátu,
- přírůstky volných míst Úřadu práce ČR v HTML nebo XML formátu,
- volná místa z jednotlivých okresů v HTML nebo XML formátu,
- seznam insolventních firem evidovaných Úřadem práce ČR v HTML nebo XML formátu,
- měsíční statistika nezaměstnanosti.

Z těchto dat by mohlo být možné například hledat závislost mezi mírou kriminality a nedostatkem „správných“ pracovních míst.

Data z portálu justice.cz

Na tomto webu je možné získat informace z prostředí soudnictví v ČR. Portál je provozován Ministerstvem spravedlnosti ČR. Je možné získat následující data:

- rejstřík trestů,
- obchodní rejstřík,
- insolvenční rejstřík,
- evidence úpadců,
- seznamy soudců, státních zástupců, soudních znalců,
- počet žádostí o výpis z trestního rejstříku podle měsíců (CSV, 2008-2015),
- počet žádostí o výpis z trestního rejstříku dle krajských úřadů (CSV, 2008-2015).

Data České správy sociálního zabezpečení

Na adrese <http://www.cssz.cz/cz/o-cssz/informace/statistiky/> lze nalézt následující data:

- důchodová statistika,
- nemocenská statistika, pracovní neschopnost (v průběhu let),
- vývoj příjmů pojištěoven v průběhu měsíců a let.

Pomocí těchto dat by bylo možné analyzovat proč a kdy se lidé více pojišťují. Dále jsou k dispozici údaje o plnění povinností zaměstnavatelů (plátců pojistného, nedoplatky, důchodové a nemocenské pojištění) rozdělené dle krajů. Záznamy o neplnění těchto povinností představují přímo kriminální jednání. Bohužel veškerá data jsou ve formátu PDF, čímž značně ztěžují možnost strojového čtení.

Data z Českého statistického úřadu

Portál <https://www.czso.cz/> je bohatý zdroj informací. Je zde možné získat data nejen v souborech formátu PDF, ale také jsou k dispozici jako HTML tabulky. Zdrojem těchto dat je veřejná databáze ČSÚ. Jako potenciálně zajímavá data byla vybrána:

- průměrné hrubé měsíční mzdy podle pohlaví v krajích,
- průměrné hrubé měsíční mzdy zaměstnanců podle hlavních tříd klasifikace zaměstnání CZ-ISCO v krajích,
- výdaje na dávky nemocenského pojištění podle druhů dávek v krajích,
- vyplacené dávky státní sociální podpory podle druhu dávek v krajích,
- vyplacené dávky státní sociální podpory podle druhu dávek v okresech,
- sociální služby poskytované v zařízeních sociálních služeb v krajích,
- průměrný počet osob nemocensky pojištěných v krajích,
- zaměstnanost/nezaměstnanost v krajích,
- počet cizinců podle okresů, věku,
- obvinění podle pohlaví,
- trestné činy a jejich objasněnost.

Rejstříky jako další zdroj informací

Na serveru <http://www.rejstriky.info/> je dostupný vyhledávač s analytickým výstupem. Portál poskytuje zajímavou službu grafického zobrazení podnikatelských vazeb mezi jednotlivými podnikatelskými subjekty. Je možné vyhledávat v následujících kategoriích:

- obchodní rejstřík,
- živnostenský rejstřík,
- insolvenční rejstřík,

- administrativní registr ekonomických subjektů (ARES),
- platiči/neplatiči DPH.

3.2 Otevřená data

Nezisková organizace Open Knowledge definuje otevřená data jako: „Taková data, která jsou zveřejněná na internetu způsobem, který neomezuje žádné uživatele ve způsobu jejich použití pro nekomerční i komerční účely (technicky ani legislativně) a opravňuje všechny uživatele k jejich dalšímu šíření, pokud tímto šířením nedojde k omezení práv ostatních uživatelů.“

Smyslem otevřených dat je tedy možnost nahlédnutí či využití informací pro odbornou veřejnost. Může tak dojít k vytvoření zajímavých aplikací či analýz veřejností, na které veřejná správa např. nemá finance. V praxi jsou otevřená data publikována orgány veřejné správy za účelem vyšší transparentnosti.

Otevřená data musí být:

- publikována v maximálním možném rozsahu a kvalitě,
- snadno dostupná a dohledatelná na internetu,
- strojově čitelná,
- zpřístupněna pod jasně danými licencemi pro komerční i nekomerční použití.

Z těchto důvodů není možné označit za otevřená data například zveřejněné tabulky ve formátu PDF, či ve formě HTML stránky. Na portálu <http://portal.gov.cz/portal/obcan/rejstriky/data/> je možné použít nástroj *Národní katalog otevřených dat (NKOD)*, který umožňuje snažší orientaci a vyhledávání v otevřených datech publikovaných veřejnou správou ČR. Otevřená data představují díky své relevanci významný zdroj informací pro budoucí predikční modelování.

3.3 Další data, která mohou přispět k predikci

Chování pachatelů může ovlivňovat mnoho aspektů. Při zkoumání zdrojů dat pro objasnění trestné činnosti, bylo zjištěno, že k odhalení může významně přispět pochopení chování pachatele. Hledat široké souvislosti je věc příslušného vzdělání,

praxe i talentu kriminalistů analytiků. Ke každému typu trestné činnosti je třeba přistupovat odlišně. Hledat zobecnění příčin, a přenést je do návrhu dat, je náročný úkol, proto bylo k tomuto řešení přistoupeno v určité míře „laicky“. V datech, která byla v této práci použita (viz. kapitola 5.7), chyběly údaje o počasí, které jinak vytvořená aplikace u nových dat automaticky získává (viz. kapitola 5.4). To by mohlo některé typy kriminálního jednání ovlivnit. Proto dalším zdrojem dat je informace o aktuálním počasí v době trestného činu. Jedna z nejvýznamějších informací je údaj o lokaci, kde k činu došlo. Vzhledem k možnosti využití nových technologií se nabízí využití systému GPS. Tato varianta byla také použita v praktické části.

3.4 Formáty dat

Kromě dat ze serveru České správy sociálního zabezpečení lze všechna data exportovat do přijatelného formátu (XML, XLS), se kterým lze dále pracovat. K datům byla vždy dostupná také metadata, která strukturu popisovala. Ve formátu XLS byly názvy sloupců popsány. Pro formát XML platilo, že lze případně stáhnout doplňující soubor, který jednotlivé identifikátory spojí se slovními popisky. Při prozkoumávání dat nebylo zjištěno nic, co by dělalo softwaru Knime problém s konverzí. Data pochází z oficiálních zdrojů, tudíž by měla být korektní a úplná.

4 Typy kriminálního jednání a datamining

Nasazení dataminingových úloh do řešení kriminálního jednání lze vidět ze dvou úhlů pohledu. Jeden přístup může ve velkých datech hledat skryté korelace, spojitosti, podobnosti napovídající, kdo by mohl být pachatelem. Podmínkou je mít data dostatečně velká. Druhý postup by analyzováním historie predikoval potřebu nasazení techniky a lidí v místě a čase tak, aby se čin nestal. Opět lze jen zopakovat potřebu mít správná data. Závěr několika jednání na městské policii, policii ČR a osobních jednání na workshopu a konferenci „Mapy budoucnosti“ potvrdil, že se postupně nové technologie zavádějí. Jsou k dispozici různé mapové materiály „hot spotů“, veřejně dostupný web <http://www.mapakriminality.cz/>, kde lze dohledat souhrnné statistiky s různým stupněm filtrování, či zobrazovat hustotu kriminality.

4.1 Trestní kriminalita a datamining

Pro prevenci kriminality lze použít data o trestných činech jako jsou vraždy, znásilnění, loupeže, rvačky, krádeže automobilů, vykrádání aut, výroba a distribuce drog, řízení pod vlivem a pracovat s daty podobně jak je popsáno v praktické části. Trestné činy, jak je kategorizuje kriminální oddělení PČR jsou uvedené v příloze 5 . Nasazením algoritmů shlukové analýzy lze nacházet multidimenzionální podobnosti či nepodobnosti popsanych trestných činů. Predikovat jejich výskyt v místě a čase a například cíleně nasadit na kritická místa hlídky. Jednotlivé kategorie trestných činů je třeba řešit vždy podle získaných dat a specifik příslušného trestního jednání. Jinak bude postavený model pro vloupání do bytových jednotek, jinak pro vraždy či jinou násilnou činnost.

4.2 Ekonomická kriminalita a datamining

Jedná se typicky o bankovní podvody, či praní špinavých peněz, které by mohlo být odhaleno dříve, než dojde k finančním ztrátám. Predikovat, zda klient bude schopen úvěr splácet, nemusí mít a ve většině případů nemá kriminální základ. Procento těch, kteří žádají o úvěr s úmyslem jej nesplácet a jedná se o úvěrový podvod, bude malé, ale škoda s tím spojená může být velká. Jinak je to s podezřelými transakcemi, které mohou banku závažně poškodit. Nasazením klasifikačních algoritmů, resp. modelů, a také algoritmů shlukové analýzy, se lze ekonomickým škodám bránit. Pojistné a dotační podvody jsou další kategorie podvodného jednání, které lze takto predikovat.

4.3 Internet a trestní jednání

Internet, kromě nesporných pozitiv, přinesl nové hrozby. O uživatelích internetu je možné zjistit mnoho dat, a o neinformovaných uživatelích „skoro vše“. Na sociálních sítích uživatelé na sebe prozrazují detaily o svých zájmech, o tom kde se právě nachází, kdy a na jak dlouho opouští domov apod. Z těchto informací a pohybu uživatele internetem lze chytře sestavit profil uživatele a cíleně s ním posléze manipulovat. Ovlivňování uživatelů a vytváření zájmových skupin s agresivním nepřátelským zaměřením, šíření radikálních teroristických myšlenek, je velmi nebezpečné. Úlohy zabývající se odhalováním nebezpečných osob tímto způsobem se již dnes v praxi používají.

5 Vytvořený systém pro predikci výskytu trestné činnosti

Při účasti na workshopu Mapy budoucnosti a následných jednáních na PČR i MP v Liberci, vznikla myšlenka doplnit analytickou část predikování výskytu trestní činnosti o mobilní aplikaci, která umožní do systému posílat data z terénu, tato data následně v kanceláři upravit a použít pro modelování. Výsledkem pro sledované trestné činy vloupání do budov, kapesní krádeže a krádeže automobilů, bude mapa „hot spotů“ a doporučení, kdy a kde je potřeba nasadit policejní hlídky nebo kamerový dohled. Modely jsou postavené na algoritmech shlukové analýzy.

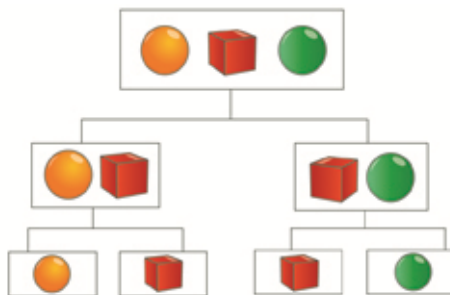
5.1 Teoretický podklad

5.1.1 Shluková analýza

Shluková analýza[1][2] je statistická metoda patřící mezi „učení bez učitele“. Cílem takovéto analýzy je sjednocení určitých objektů do shluků, tak aby si objekty v jednom shluku byly navzájem podobné a zároveň byly odlišné od objektů mimo tento shluk. Shlukovou analýzu je možné provádět na množině objektů, které jsou popsány prostřednictvím stejného souboru znaků, či na množině znaků, které jsou charakterizovány pomocí skupiny objektů, nesoucí tyto znaky. Shlukovací metody lze rozdělit podle cíle na:

- **Hierarchické shlukování:** Je systém navzájem různých neprázdných podmnožin množiny X , v němž průnikem každých dvou podmnožin je buď jedna z nich nebo prázdná množina a v němž existuje alespoň jedna dvojice podmnožin, jejichž průnikem je jedna z nich. V průběhu hierarchického shlukování vzniká tzv. dendrogram. Tento diagram znázorňuje jednotlivé kroky analýzy, přičemž na horizontální ose jsou vzdálenosti mezi jednotlivými shluky

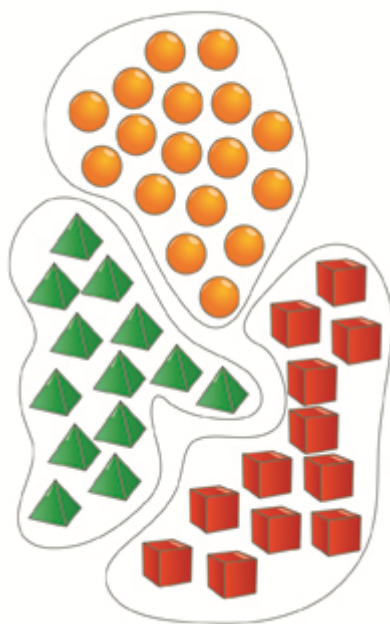
a na svislé ose jednotlivé kroky rozdělení. Volbou konkrétní úrovně lze pak vybrat optimální řešení analýzy.



Obrázek 5.1: Grafické znázornění hierarchického shlukování

Zdroj: <http://analyticstraining.com/>

- **Nehierarchické shlukování:** Je systém navzájem různých neprázdných podmnožin množiny X , v němž průnikem každých dvou podmnožin není žádná z nich.



Obrázek 5.2: Grafické znázornění nehierarchického shlukování

Zdroj: <http://analyticstraining.com/>

Znaky

Každý objekt je popsán p -ticí stavů předem daných p znaků. To znamená, že konkrétní objekt v analýze je reprezentován p -rozměrným vektorem. Znaky objektů a množiny jejich stavů mohou být:

- **Kvalitativní znaky:** Konečná množina popisujících termínů. Mohou být buď nominální nebo ordinální. Nominální termíny jsou takové, které nelze nějakým kritériem řadit, jako například barva (červená, žlutá, apod.). Ordinální termíny je možné uspořádat (malé, střední, velké). Kvalitativním znakem může být také binární znak určující pouze přítomnost daného příznaku např: savec - (ano/ne), či pohlaví - (muž/žena).
- **Kvantitativní znaky:** Jsou vyjádřeny pomocí čísla z určitého intervalu. Tato čísla mohou být celá i reálná. Např: výška - (145,6 cm)[2]

Standardizace dat

V popisných datech, což jsou znaky objektů, se může stát, že některá vstupní proměnná (znak) je řádově na vyšší škále oproti jinému znaku. Například cena auta, oproti počtu dětí, když objekt bude klient banky, o kterém zjišťujeme predikci splácení poskytnutého úvěru. Při měření „vzdáleností“ objektů nebo podobnosti objektů, by dominující znak zastínil znaky měřené na nižší škále. Proto je vhodné všechny znaky transformovat na stejnou škálu - rozsah. Pro souměřitelnost všech znaků se používá standardizace dat. Příkladem standardizace dat může být transformace na z-skóry. [2]

1. Výpočet střední hodnoty a směrodatné odchylky:

$$\bar{z}_j = \frac{1}{n} \sum_{i=1}^n z_{ij}, s_j = \left[\frac{1}{n} \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2 \right]^{\frac{1}{2}}$$

2. Přepočítání původních hodnot znaku na standardizované hodnoty:

$$x_{ij} = \frac{z_{ij} - \bar{z}_j}{s_j}$$

Dalšími metodami standardizace jsou transformace na interval $\langle 0, 1 \rangle$ nebo $\langle -1, 1 \rangle$ nebo logistická funkce, ArcTg, či jiné speciální transformace.

5.1.2 Koeficienty asociace objektů

[3] Koeficienty asociace vyjadřují míru podobnosti jednotlivých objektů za použití všech atributů. Tento způsob zjišťování podobnosti, se tedy týká objektů, reprezentovaných pouze kvantitativními znaky. K vizualizaci pomáhá asociační tabulka.

Tabulka 5.1: Asociační tabulka

		xi	
		1	0
xj	1	a	b
	0	c	d

Ve výše zobrazené tabulce je možné vidět dva objekty X^i a X^j , které obsahují p atributů. Písmena v tabulce představují:

- a – počet znaků, kde oba objekty dosahují hodnoty 1,
- b – počet znaků, kde objekt X^i má hodnotu 0 a objekt X^j má hodnotu 1
- c – počet znaků, kde objekt X^i má hodnotu 1 a objekt X^j má hodnotu 0
- d – počet znaků, kde oba objekty dosahují hodnoty 0.

S těmito hodnotami dále pracují konkrétní koeficienty asociace. Příklad Hamanova koeficientu asociace:

$$\frac{a + d - (b + c)}{a + b + c + d}$$

5.1.3 Metriky

[3] Geometrický model je do posuzování vnesen používáním metrik. Je-li objekt X^i , mající p znaků, tak ho lze zobrazit do p -rozměrného metrického prostoru. Metrický prostor je tvořen dvojicí (M, ρ) , kdy M , je libovolná neprázdná množina a ρ , označované jako metrika, je zobrazení: $M \times M \rightarrow \mathbb{R}$, přičemž pro objekty A, B a C z množiny M musí platit:

1. $\rho(A, B) = 0 \Leftrightarrow A = B$
2. $\rho(A, B) \geq 0$
3. $\rho(A, B) = \rho(B, A)$

$$4. \rho(A, C) \leq \rho(A, B) + \rho(B, C)$$

Příklady metrik:

- Euklidovská: $d_e(A, B) = \sqrt{\sum_{i=1}^p (a_i - b_i)^2}$
- Manhattanská: $d_m(A, B) = \sum_{i=1}^p |a_i - b_i|$
- Chebyshevova: $d_{ch}(A, B) = \max |a_i - b_i|$

5.1.4 Algoritmus k-means

K-means, neboli k-středů, je metoda nehierarchické shlukové analýzy popsaná v 1967 Johnem MacQueenem. Úkolem algoritmu je rozdělit množinu vektorů dimenze p do k podmnožin (shluků) tak, aby byla nejmenší suma vzdáleností jednotlivých vektorů od středu příslušné podmnožiny. Celý algoritmus lze implementovat pomocí následujících 4 kroků:

1. Rozdělení všech objektů do k neprázdných shluků.
2. Výpočet těžiště každého shluku při tomto rozdělení.
3. Rozřazení všech objektů do shluků, podle nejmenší vzdálenosti k těžišti.
4. Pokud došlo ke změně přiřazení, opakuj od kroku 2.

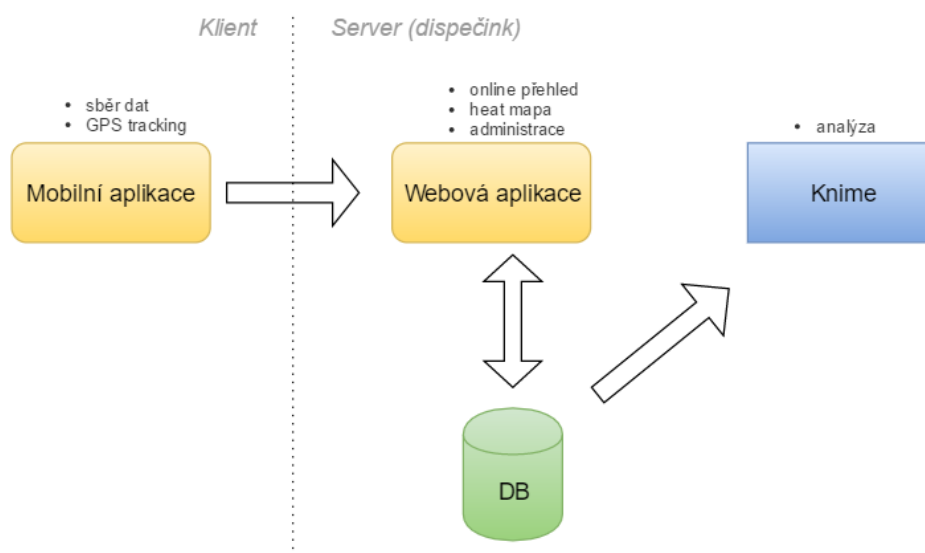
Tento postup je v praxi velmi citlivý na počáteční pozice jednotlivých shluků. Jsou-li tato inicializační těžiště vybírána náhodně, může dojít k odlišným výsledkům analýzy. Je proto třeba pro hledání optimálního výsledku algoritmus několikrát zopakovat. Tomuto jevu je možné se vyhnout pomocí apriorní znalosti o vstupních datech a nevolit tak počátky zcela náhodně.

Vlastnosti algoritmu k-means:

- jednoduchost,
- prvky se mohou přeskupovat mezi shluky,
- pouze pro kvantitativní data,
- konverguje v konečném počtu kroků k nějakému řešení,
- může existovat více řešení v závislosti na počátečních podmínkách.

5.1.5 Popis systému

V rámci diplomové práce byl vytvořen systém, který umožňuje řešit možnou úlohu prevence proti kriminalitě. Inspirací pro tento systém byl americký program Blue CRUSH. Cílem bylo vyřešení celého řetězce počínaje od sběru dat, po jejich vyhodnocení. Celý systém se skládá ze tří základních částí, přičemž každá část je autonomní nástroj. První část systému je mobilní aplikace, kterou používá policista v terénu pro sběr dat v reálném čase. Tato data jsou následně odeslána na webový server, který představuje druhou část systému, kde se data zpracují a uloží do databáze. Webová aplikace také mimo jiné slouží jako dispečink a administrační rozhraní. Poslední částí je analytický nástroj Knime, který se připojuje do databáze a pomocí vytvořených modelů analyzuje data a produkuje výstup.



Obrázek 5.3: Diagram vytvořeného systému

5.2 Mobilní aplikace

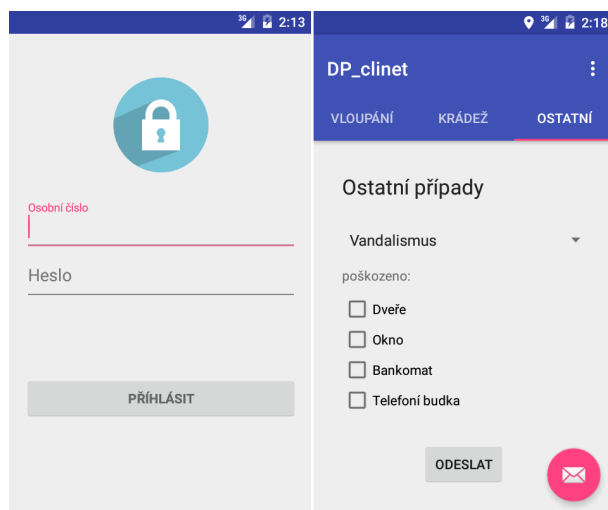
Jedna z hlavních motivací ohledně získávání dat byla rychlost. Právě možnost sledovat situaci v reálném čase otevírá další možnosti pro zefektivnění práce. Proto byla zvolena možnost využití mobilního zařízení, které by měla hlídka přímo v terénu a odesílala veškeré informace pomocí internetu na stanici. Jako platforma byl zvolen operační systém Android. Důvodem tohoto rozhodnutí byla především jistá zkušenost s programováním pro tuto platformu. Základní funkce mobilní apli-

kace jsou GPS tracking a v případě vytvoření nového záznamu o trestné činnosti pak odeslání těchto dat.

Po spuštění uživatel vidí přihlašovací formulář, kde je potřeba vyplnit služební číslo policisty a jeho heslo. Po odeslání dojde k autentizaci na straně webového serveru. Po úspěšném přihlášení je uživatel vyzván aby uvedl, zda se jedná o pěší hlídku, či hlídku s vozem. Následuje přechod na základní obrazovku současně se spuštěním GPS trackingu. Aktuální pozice zařízení se odesílá každých deset vteřin společně s příslušným časem, informací o přihlášeném uživateli a pracovním módem (vůz, pěší). Aplikace se skládá ze tří fragmentů (obrazovek), přičemž každý představuje typ trestné činnosti. Na každém fragmentu je dále možné upřesnit činnosti, či vyplnit informace o okolnostech. Mezi těmito fragmenty je možné přepínat pomocí „slidu“ prstem vpravo či vlevo, nebo pomocí navigace v horní části aplikace. V každém fragmentu je možnost odeslat záznam o trestném činu pomocí tlačítka v dolní části. Každý odeslaný záznam musí mít údaj o poloze, proto je nutné vždy po spuštění počkat na vyhledání satelitů GPS. Jakmile již aplikace získá tento údaj je možné kdykoliv odeslat záznam o trestné činnosti. V případě, že v místě odeslání není signál družic dostatečný, použije se vždy poslední známý údaj o poloze. Stáří poslední informace polohy je možné zkontrolovat vždy v potvrzovacím okně, které se vyvolá při odeslání každého záznamu o trestné činnosti.

Požadavky na mobilní aplikaci jsou:

- Android 4.1 a novější,
- GPS modul,
- přístup k internetu.

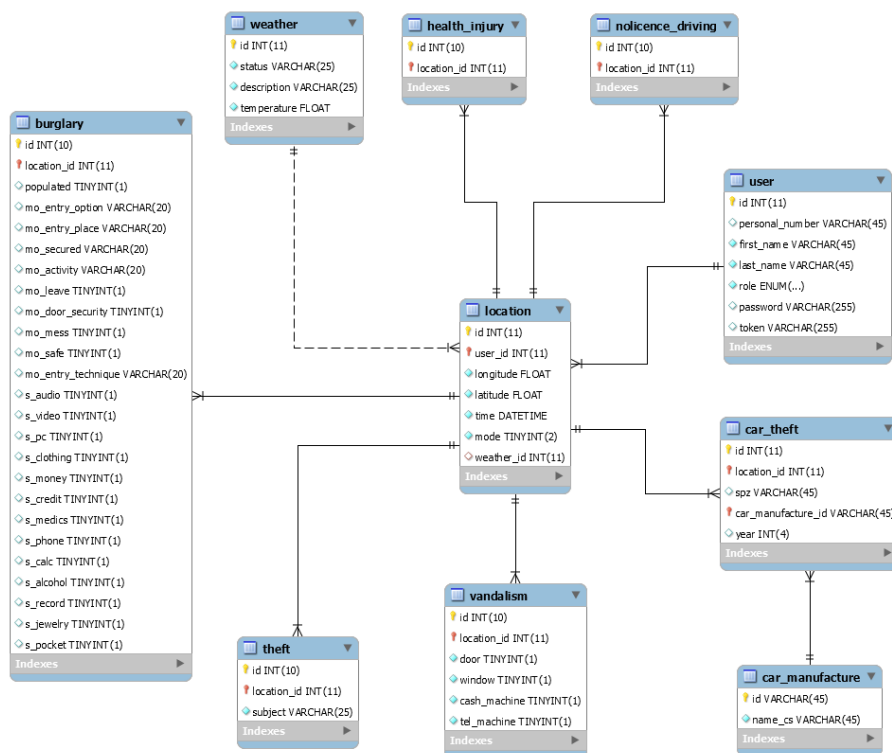


Obrázek 5.4: Interface mobilní aplikace

5.3 Databáze

Pro uchování dat byl zvolen databázový systém MySQL. Důvodem byla použitelnost ve spojení s webovou aplikací a zároveň podpora ze strany nástroje Knime. Návrh databáze bylo třeba realizovat podle použitých dat. Původním záměrem bylo získat konkrétní podobu dat pomocí spolupráce s policií ČR v Liberci. Ačkoliv se předpokládalo, že pro účely diplomové práce bude možné získat tato anonymizovaná data, nakonec k tomu bohužel nedošlo. Byl proto použit cvičný dataset, jehož struktura byla s policií konzultována. Po konzultaci bylo zjištěno, že tato datová struktura je „podobná“ té, kterou používá Policie ČR a může tak posloužit pro návrh databáze.

Použitý dataset měl podobu jednoho CSV souboru, kde byly zapsány veškeré trestné činy jako jednotlivé řádky. Při návrhu bylo třeba tabulku normalizovat a následně upravit pro vytvořenou aplikaci.



Obrázek 5.5: Schéma navržené databáze

Ze schématu je patrné že veškeré události se odkazují na tabulku *location*. Záznamy v této tabulce představují jak data získaná GPS trackingem tak polohu trestných činů. Každý řádek také nese informaci o uživateli, módu ve kterém hlídka operovala, popř. počasí v dané lokaci v příslušném čase.

5.4 Webová aplikace

Pro vyřizování požadavků z mobilní aplikace a správu databáze byla vytvořena webová aplikace. Při tvorbě byl použit český webový PHP framework *Nette*.

Důvody použití *Nette frameworku*:

- kvalitní objektový návrh MVC
- vysoké nároky na bezpečnost
- výborné ladící nástroje (*TRACY*)
- česká komunita

Zpracování požadavků

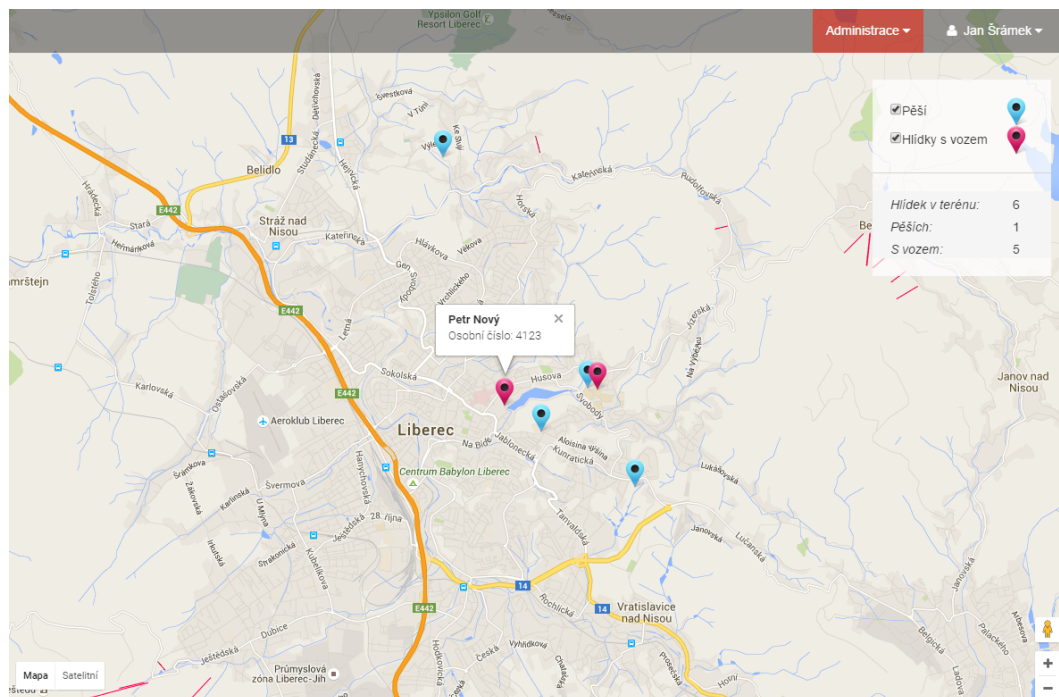
Prvním úkolem webové aplikace bylo zpracování požadavků z mobilní aplikace. Tyto požadavky mají formu HTTPS GET požadavků. Parametry těchto požadavků představují zasláná data. Pro zachování relevance dat bylo třeba dodržet pravidlo, že v jednom okamžiku, může být pod jedním uživatelským účtem přihlášeno vždy pouze jedno zařízení. Tento problém byl v systému vyřešen pomocí *tokenů*. Princip je takový, že při každém novém přihlášení z mobilní aplikace, se na straně serveru vygeneruje unikátní hash řetězec a uloží se do databáze k příslušnému uživateli. V tento okamžik je také zaslán do mobilní aplikace a jeho přijetí slouží zároveň jako potvrzení přihlášení. Tento řetězec je generován pomocí funkce `bcrypt` z `id` uživatele a je k němu připojena náhodná sůl. Tímto postupem by se mělo zamezit možným kolizím tokenů. Token je zasílán v každém požadavku a na jeho základě dochází k identifikaci uživatele. V případě, že token při zpracování požadavku není v databázi nalezen, je jisté že byl vygenerován již nový, v důsledku přihlášení z jiného zařízení. V tomto okamžiku je na mobilní zařízení zaslán příkaz k odhlášení uživatele. Důsledkem je vždy priorita nového přihlášení.

Při zpracování požadavků přidávajících nový záznam trestné činnosti, aplikace získává data o počasí. Zdrojem dat je portál <http://openweathermap.org>, který poskytuje komunikační API pro získávání meteorologických dat na základě GPS souřadnic. Služba nabízí 60 požadavků za minutu zdarma, což bylo pro potřeby diplomové práce zcela dostačující. V získaných datech je vždy uveden popis počasí, bližší informace a teplota. Všechny tyto informace jsou vždy ukládány do databáze.

On-line dispečink

Webová aplikace také umožňuje monitorovat aktuální situaci v terénu. V levém navigačním panelu jsou položky „Přehled“ a „Heat mapa“. Přehled nabízí on-line sledování všech aktivních hlídek v terénu. Tato funkce byla realizována pomocí Google Maps API. Jelikož se data v diplomové práci soustředí na Liberec a jeho blízké okolí, je vždy při načtení stránky mapa zaměřena na tuto lokalitu. Na mapové podklady jsou vykreslovány markery označující poslední zaznamenanou polohu daného uživatele. Markery mají dvě možné barvy, přičemž modrá barva znamená pěší hlídku a červená barva hlídku, která má k dispozici vůz. Jestliže mobilní zařízení neodesílá pozici déle než tři minuty, příslušný marker zešediví. Na ikony je možné kliknout, čímž se rozbálí informační okno o dané hlídce. V tomto okně je vypsáno jméno

uživatelé a služební číslo policisty. Merkery jsou aktualizovány pravidelně každých deset vteřin. Načítání i aktualizace pozic bylo vytvořeno pomocí AJAXU a probíhá tak na pozadí.



Obrázek 5.6: Ukázka přehledu hlídek v terénu

Heat mapa nabízí možnost zobrazení kriminálních činů pomocí dvourozměrné tepelné mapy. Tato mapa je opět vytvořena pomocí Google Maps API. V pravé části mapy se nachází filter. Filter slouží k možnosti zobrazení jen vybraných druhů trestných činů. Dále je možné vybrat pomocí datepickerů konkrétní časové období. Implicitně je nastaveno období za poslední rok. Toto zobrazení slouží především pro nalezení nejvíce rizikových oblastí, tzv. „hot-spotů“.

Administrace

V aplikaci byla vytvořena víceúrovňová struktura oprávnění. Každý uživatel je zařazen do určité skupiny, které náleží různá oprávnění. Nadřazená skupina vždy rozšiřuje skupinu podřízenou, čímž přebírá všechny její vlastnosti. Vytvořené skupiny v aplikaci jsou:

- guest
- user

- admin

Skupina *guest* je v aplikaci prakticky nevyužita, představuje spíše možnost pro případné restriktce klasických uživatelů. Uživatelé skupiny *user* se mohou přihlásit do webové aplikace a mají k dispozici on-line dipečink, tedy přehled a heat mapy. Pokud se přihlásí uživatel ze skupiny *admin*, zobrazí se v pravé horní části navigace, vedle uživatelského jména, položka „Administrace“. V administraci je možné spravovat uživatelské účty a veškeré trestné činy v databázi. Uživatelé je možné v administraci vyhledat podle jména, příjmení, či služebního čísla. Během zadávání znaků do vyhledávacího formuláře systém našeptává možné výsledky. V sekci administrace vloupání je možné, mimo editace, také záznamy zpracovávat. Zpracováním je myšleno doplnění informací, které se neodesílají přímo z mobilní aplikace. Pro práci s databází bylo v aplikaci použito databázové ORM Lean Mapper. Toto rozšíření frameworku umožňuje práci s databází na úrovni objektů (entit), které si uživatel nadefinuje. Jednou ze zajímavých vlastností je možnost entitě, která reprezentuje záznam v databázi, jako vlastnost přidat kompletní jinou entitu či jejich kolekci.

5.5 Knime

Poslední částí systému jsou vytvořené modely, které analyzují získaná data. K vytvoření modelů bylo použito analytické prostředí Knime. Jedná se o komerční open-source nástroj vyvíjený ve spolupráci s německou univerzitou v Kostnici. Aplikace je založena na licenci GPLv3. Knime nabízí editor, ve kterém je uživateli umožněno pomocí grafického rozhraní vytvořit „procesní datový tok“. Ten se skládá z předpřipravených uzlů, které uživatel spojí a vytvoří tak řetězec, který je při spuštění sekvenčně zpracován. V aplikaci je k dispozici mnoho uzlů nabízejících funkce pro přípravu dat, klasifikaci, strojové učení, či tvoření grafických výstupů. Uživateli je také umožněno, v případě absence uzlu s požadovanou funkcionalitou, vytvořit vlastní uzel, jehož logika je naprogramována pomocí jazyka JAVA, R a dalších. Vytvořené modely v rámci diplomové práce jsou popsány v kapitole 5.8 a 5.9.

5.6 Bezpečnost

Vzhledem k povaze dat bylo nutné zohlednit bezpečnost vytvořeného systému. Z celkového návrhu, potažmo směru toku dat, je zřejmé, že mobilní aplikace nemá přístup k databázi. Nicméně pro přihlášení do mobilního zařízení je třeba zaslat přihlašovací údaje směrem k serveru. Pro minimalizaci rizik útoku typu „man in the middle“ je tato komunikace realizována pomocí protokolu HTTPS. Z těchto důvodů je webová aplikace hlavní částí, skrze kterou by bylo možné útočit na získaná data.

Jednou z příčin volby Nette byl důraz na bezpečnost v tomto frameworku. Framework sám o sobě totiž kvalitně ošetřuje veškeré vstupy formulářů. Mezi typy útoků, které framework úspěšně řeší, patří:

- cross-site scripting (XSS),
- cross-site request forgery (CSRF),
- URL attack, control codes, invalid UTF-8
- session hijacking, session stealing, session fixation

Nette framework také nabízí několik způsobů, jak řešit oprávnění uživatelů. Ve vytvořené aplikaci bylo použito statické ACL (Access Control List). Tímto způsobem bylo kontrolováno při zpracování všech formulářů a před vykreslením jednotlivých stránek, zda je uživatel přihlášený a jestli má pro danou akci příslušná oprávnění, čímž v kombinaci s kvalitně ošetřenými vstupny proti potenciálním útokům, je výrazně nižší možnost napadení. Dále, jak již bylo zmíněno v kapitole 5.2, celá aplikace také pracuje na zabezpečeném protokolu HTTPS.

5.7 Použitá data

Pro praktické úlohy bylo potřeba pracovat s větším množstvím relevantních dat. Protože se nepodařilo získat skutečná policejní data, bylo potřeba vytvořit vlastní množinu. Pro úlohy 5.8 a 5.9 byla použita „školní data“, která mají podobnou strukturu i obsah jako data skutečná, ale poloha byla z jiné lokality.

Na základě těchto dat došlo také k návrhu databáze. Tento dataset měl podobu jednoho souboru ve formátu csv. Každý řádek tohoto souboru představoval konkrétní záznam o trestné události a sloupce popisovaly okolnosti případu.

Tabulka 5.2: Struktura a popis vstupního souboru

Atribut	Popis
cislo_pripadu	Identifikační číslo případu
Kod	Kódové značení události
souradnice_X	X-ová souřadnice na mapě
souradnice_Y	Y-ová souřadnice na mapě
datum	Datum, ve kterém došlo k případu
MO_vstup	Způsob vstupu
MO_misto	Místo vstupu
MO_zabezp	Způsob zabezpečení
MO_alarm	Alarm
MO_cinnost	Zaměření budovy
MO_odchod	Opuštění budovy
MO_zabdvere	Zabezpečení dveří
MO_neporadek	Zanechán nepořádek
MO_sejf	Pokus o vyloupení sejfu
MO_zpusobvstupu	Okolnosti vstupu
Kaud	Odcizený audio přehrávač
Kvid	Odcizený video přehrávač
Kpocitac	Odcizený počítač
Pobleceni	Ukradeno oblečení
Ppenize	Ukradeny peníze
Pkredit	Ukradena kreditní karta
Plek	Ukradeny léky
Ptelefon	Ukraden telefon
Phodiny	Ukradeny hodiny
Pkalk	Ukradeny kalkulačky
Palkoh	Ukraden alkohol
Pzaznam	Ukraden audio záznam
Psperky	Ukradeny šperky
Ppeněz	Ukradena peněženka
Pdvere	Ukradeny dveře
Pokno	Ukradeno okno
Pautomat	Vykradený výdejní automat
Ptelauto	Vykradený telefonní automat

Byl proto vytvořen script pro migraci těchto dat do databáze vytvořené aplikace. Komplikaci představovaly také údaje o poloze, kde k trestné činnosti došlo. V datasetu byly nereálné X,Y souřadnice které bylo potřeba přepočítat na GPS koordináty města Liberec. Údaje o poloze v datasetu tvořily také při větším přiblížení přesnou čtvercovou síť. Pro důvěrnější podobu dat byl každé nové poloze přičten náhodný koeficient. Ukázka vzorce pro přepočet polohy:

```
'longitude' => (  
  ($value - $this::X_KNIME_MIN) / $this::X_KNIME_RANGE)  
  * $this::X_GPS_RANGE  
  + $this::X_MIN  
  + rand(($this::X_GPS_RANGE/100) / 50, ($this::X_GPS_RANGE/100)  
)
```

Všechny záznamy, které byly vloženy migračním scriptem, mají nastaveného uživatele $id = 5$, což je uživatel pod kterým se nelze do aplikace přihlásit a slouží pouze pro účely migrace.

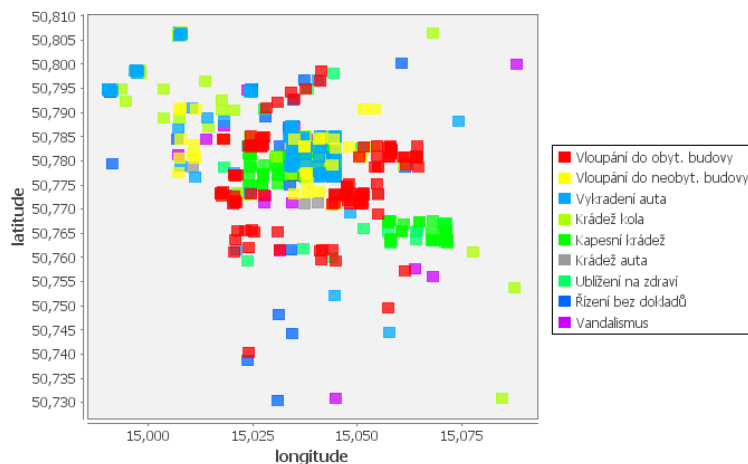
Pro úlohu 5.10 byla nasimulována vstupní data ve spolupráci s Městskou policií v Liberci. Jedná se záznamy odcizených vozidel v období dvou měsíců. Tato data byla shromážděna pomocí vytvořené mobilní aplikace a nebylo je tak potřeba upravovat. Ze stejného důvodu již také v datech nechyběla informace o počasí.

Popis úlohy

Cílem praktické úlohy bylo vytvořit softwarový model řešení pro odhalování spojitostí mezi trestnými činy a následný návrh zefektivnění práce policejních složek. Prvním úkolem bylo nalezení případů vloupání do budov, které vykazují podobné charakteristiky a následně stanovit rozložení policejních hlídek tak, aby vhodně pokrývaly oblasti s výskytem vloupání podle denní doby. Druhým úkolem bylo nalezení oblastí, ve kterých dochází ke kapesním krádežím v souvislosti s ročním obdobím. Následně pak identifikovat časové úseky a místa, kde byl abnormální nárůst právě těchto trestných činů. Pro vytvoření návrhu byla zvolena aplikace pro statistické zpracování dat Knime. Prostředí Knime dokáže pracovat s typem datetime a umožňuje tak snadno odvodit konkrétní hodnoty časových jednotek. Pro tyto úlohy byly dovozeny sloupce *hodina* a *týden*.

Pro zadání týkající se vloupání do budov byla většina sloupců ponechána, aby bylo možné pomocí společných znaků později hledat spojitosti mezi jednotlivými

případy. V případě kapesních krádeží byly migrovány pouze sloupce určující polohu a časový údaj události tak, jak jej zaznamenává mobilní aplikace.



Obrázek 5.7: Všechny trestné činy

5.8 Úloha Vloupaní

V této části úlohy byla provedena analýza vloupání do budov. Cílem bylo vytvořit model, který by rozdělil jednotlivá vloupání do skupin na základě jejich polohy a času kdy se čin odehrál, tak aby toto rozdělení mohlo být použito pro efektivní rozmístění policejních hlídek. Během dne se také počet policejních hlídek mění a proto bylo zvoleno následující rozdělení dne na:

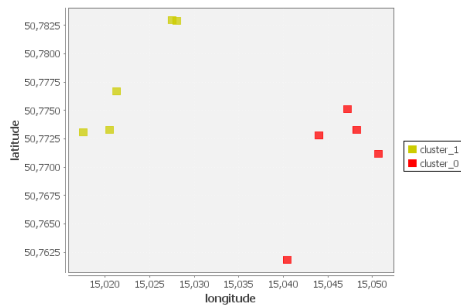
- **Ráno** - čas: 2h vč. - 7h. V tuto dobu má policie k dispozici dvě hlídky.
- **Den** - čas: 7h vč. - 18h. Přes den mohou být v terénu hlídky čtyři.
- **Večer a Noc** - čas: 18h vč. - 2h. Přes noc jsou v terénu dvě policejní hlídky.

Na základě těchto pravidel byl v datech odvozen nový sloupec *Usek_dne*, určující do jaké části dne případ patří. Pomocí uzlu CrossTab bylo možné jednoduše zjistit základní statistické údaje tohoto rozdělení:

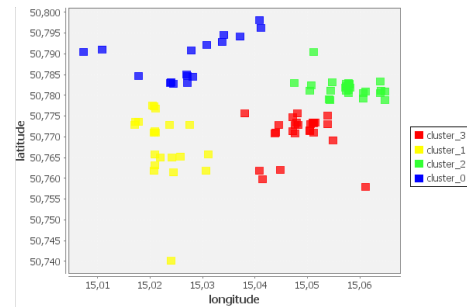
Frequency Percent	Vloupání do neobydlené budovy	Vloupání do obydlené budovy	Total
den	5	85	90
	3,9062	66,4062	70,3125
rano		10	10
		7,8125	7,8125
vecer a noc	24	4	28
	18,75	3,125	21,875
Total	29	99	128
	22,6562	77,3438	100

Obrázek 5.8: Rozdělení případů podle časových úseků

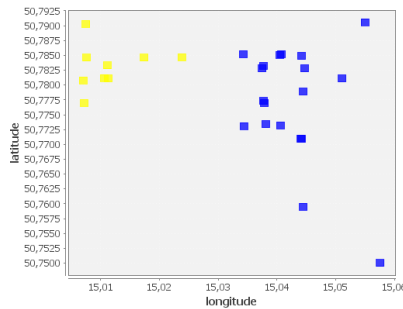
Dalším krokem byla shluková analýza dat na skupiny odpovídající počtu hlídek, které jsou v dané části dne k dispozici. Pro clusterování byl zvolen algoritmus K-means. Jelikož cílem bylo data rozdělit podle polohy, vstupními atributy byly GPS koordináty *longitude* a *latitude*. Počet clusterů odpovídal vždy počtu hlídek. V prostředí Knime je také možné nastavit maximální počet iterací algoritmu. V tomto případě bylo ponecháno přednastavených 99. Výsledky shlukové analýzy:



Obrázek 5.9: Ráno



Obrázek 5.10: Den



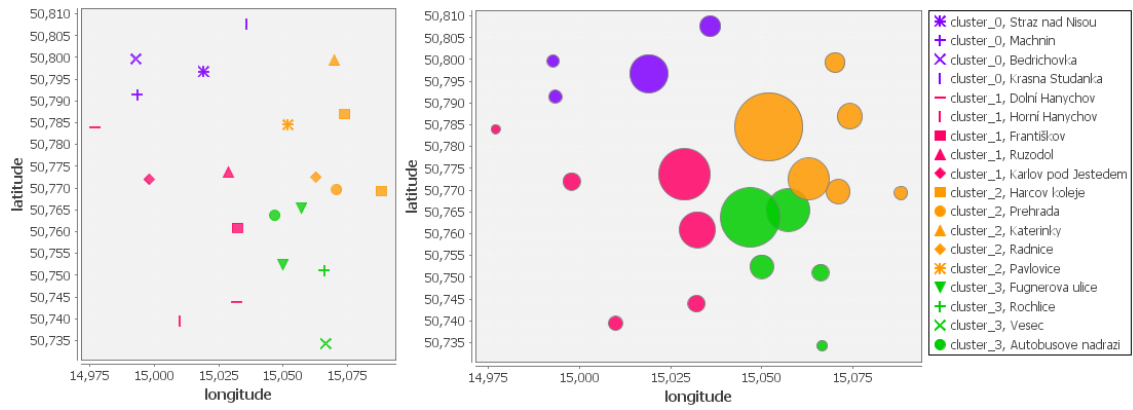
Obrázek 5.11: Večer a noc

Nasazení hlídek

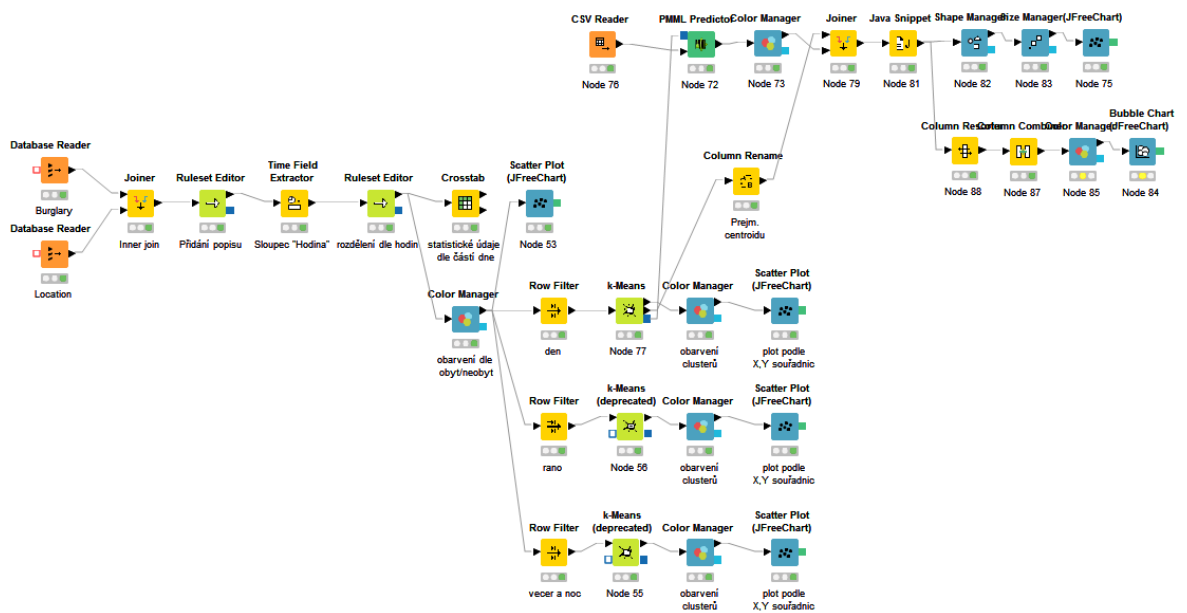
Nyní bylo potřeba jednotlivé hlídky rozmístit do terénu. Pro přiřazování oblastí ve městě byl vytvořen seznam orientačních bodů a příslušných GPS souřadnic v podobě testového souboru ve formátu CSV. Každý bod reprezentoval určitou oblast města Liberec. Na tyto orientační body byl aplikován vytvořený clusterovací model. V tuto chvíli byly body rozděleny hlídkám. Dále bylo možné využít vzdálenosti každého orientačního bodu od centroidu příslušného shluku. Tato vzdálenost resp. její obrácená hodnota reprezentuje míru závažnosti dané oblasti. K ním pro tento výpočet nemá žádný uzel, nicméně nabízí možnost naprogramování vlastního uzlu v jazyce java. V datech byl tedy vytvořen nový sloupec pomocí vzorce:

$$\frac{1}{\sqrt{(lat - lat_c)^2 + (long - long_c)^2}}$$

Rozdělení čtyř hlídek ve dne zaměřující se na vzloupání do bytů vypadá následovně:



Proud pro analýzu vloupání a následné nasazení hlídek v prostředí Knime vypadá takto:



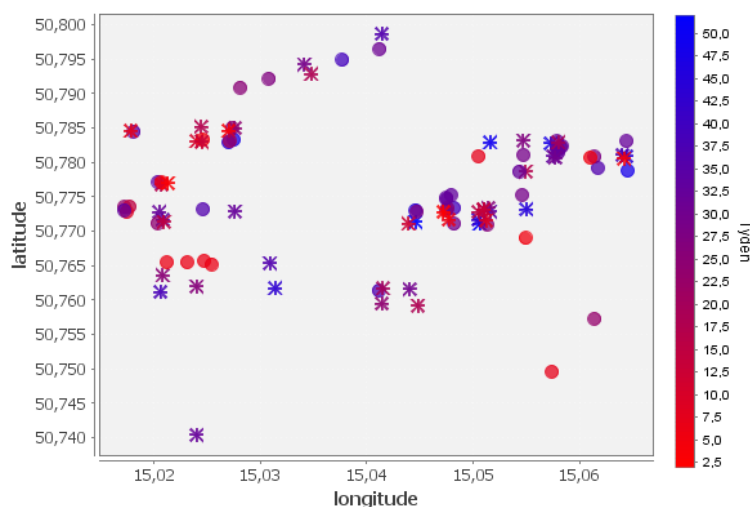
Obrázek 5.12: Ukázka proudu pro analýzu vloupání a nasazení hlídek

Nalezení souvisejích vloupání

Pomocí seskupovacích metod bylo cílem nalézt vloupání, která spolu souvisejí, na základě modus operandi (charakteristice provedení zločinu) a výpisu odcizených předmětů. Aby bylo možné na tyto vstupní atributy použít shlukovou analýzu, bylo nejprve potřeba z nominálních znaků vytvořit pomocné indikátorové atributy popisující výskyt jednotlivých příznaků, které budou metrické a mohou sloužit pro výpočet vzdálenosti. Tuto funkci v prostředí Knime provádí uzel *One to many*. S takto připravenými daty bylo možné začít s analýzou. Pro třídění byl opět vybrán algoritmus K-means. V reálné situaci by bylo možné volit počet clusterů následovně:

- Použitím algoritmu pro získání ideálního počtu skupin. Např: metoda Two-step.
- Pomocí apriori znalosti o úloze. Např: Skutečnost, že ve městě existují dvě organizované skupiny zabývající se touto trestnou činností.
- Volba náhodného počtu a následné zkoumání výsledků.

Pro tuto analýzu byly zvoleny dvě skupiny. Výsledek analýzy:

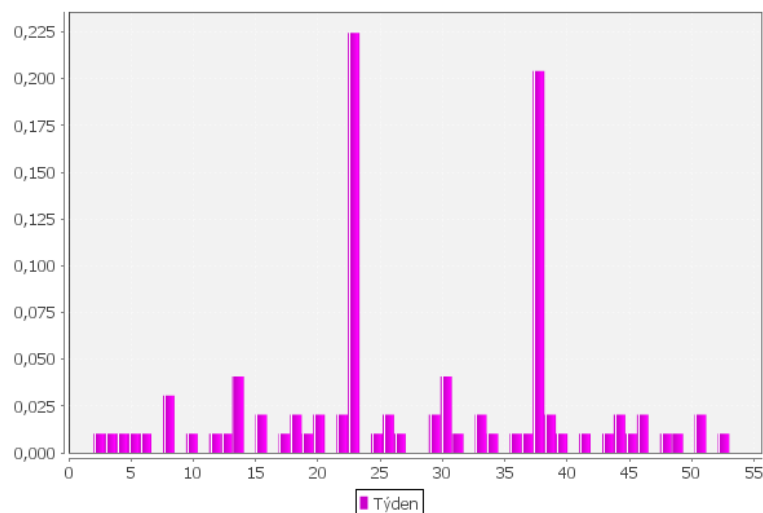


Obrázek 5.13: Výsledek analýzy souvisejících vloupání

Na výsledném grafu je první skupina označena hvězdičkou a druhá skupina kolečkem. Dále jsou všechny body obarveny dle proměnné *týden*, která do analýzy nevstupovala. Je tak možné vyzorovat nejen rozdělení podle charakteristiky zločinu, ale i časového horizontu. Návrh řešení v programu Knime:

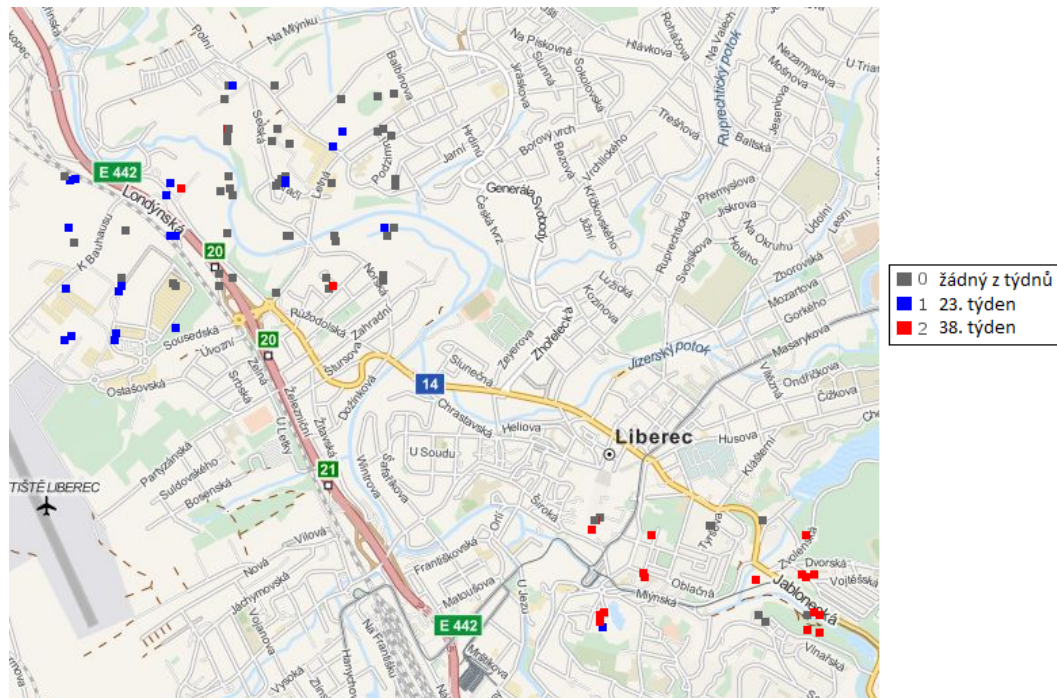
5.9 Úloha kapesní krádeže

V této úloze byla vstupem množina záznamů o kapesních krádežích. Při prohlédnutí dat pomocí histogramu byla zjištěna dvě kritická období těchto trestných činů.



Obrázek 5.14: Histogram znázorňující množství kapesních krádeží podle týdnů v roce

Z histogramu lze vyčíst, že neobvyklé množství krádeží se odehrálo ve 23. a 38. týdnu. Pomocí uzlu *Ruleset Editor* byl vytvořen nový atribut stav, který rozdělil krádeže na ty, které se staly v kritických obdobích, a mimo ně. V tuto chvíli bylo žádoucí zjistit, ve kterých oblastech ke krádežím došlo v souvislosti s tímto rozdělením.



Obrázek 5.15: Zobrazení krádeží rozdělých dle kritických období

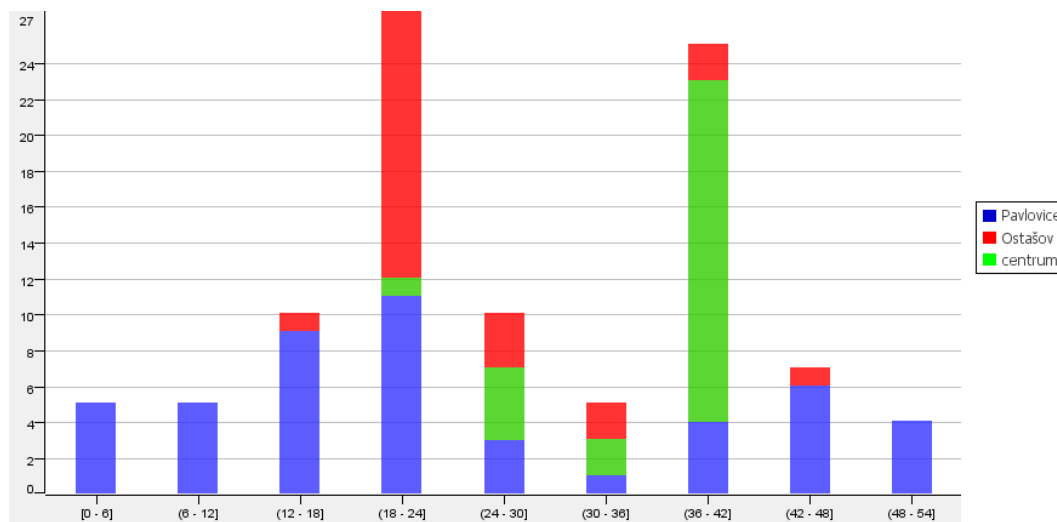
Na obrázku jsou šedě označené události, které nejsou ani v jednom kritickém týdnu. Červeně jsou označeny krádeže v týdnu 23. a modře ve 38. Na první pohled lze vypožorovat, že obě kritická období jsou také charakteristická svojí polohou. Výskyt krádeží tak bylo možné rozdělit na tři oblasti odpovídající souřadnicím Ostašova, Pavlovic a centra. Pro toto rozdělení byl vytvořen atribut *Oblast*. Pravidla pro rozdělení do konkrétních oblastí byla odvozena ze zobrazené mapy takto:

```
$longitude$ < 15.03373712 => "Ostasov"
```

```
$latitude$ < 50.768930233333 => "centrum"
```

```
$longitude$ > 15.03373712 AND $latitude$ > 50.768930233333 => "Pavlovic"
```

Dále bylo možné pozorovat krádeže v týdnech, které již byly rozděleny dle oblastí.

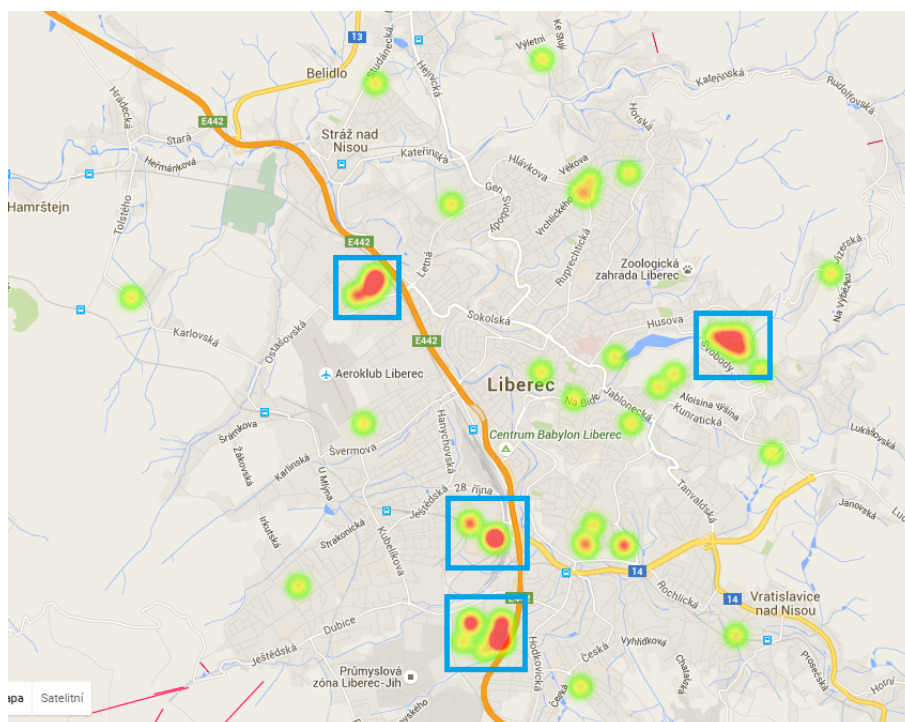


Obrázek 5.16: Histogram krádeží obarvený podle oblastí

Z obrázku lze zjistit, že v oblasti centra a Ostašova se vyskytují kapesní krádeže pouze v polovině roku, zatímco v Pavlovicích k nim dochází během celého roku.

5.10 Úloha krádeže vozidel

Cílem této úlohy bylo najít a analyzovat kritické oblasti, ve kterých dochází ke krádežím automobilů. Pro nalezení těchto míst (hot-spotů), byl použit výstup z heat mapy, která je součástí vytvořené webové aplikace. Zobrazení krádeží vozidel na tepelné mapě vypadá následovně:



Obrázek 5.17: Zobrazení rizikových oblastí krádeží automobilů

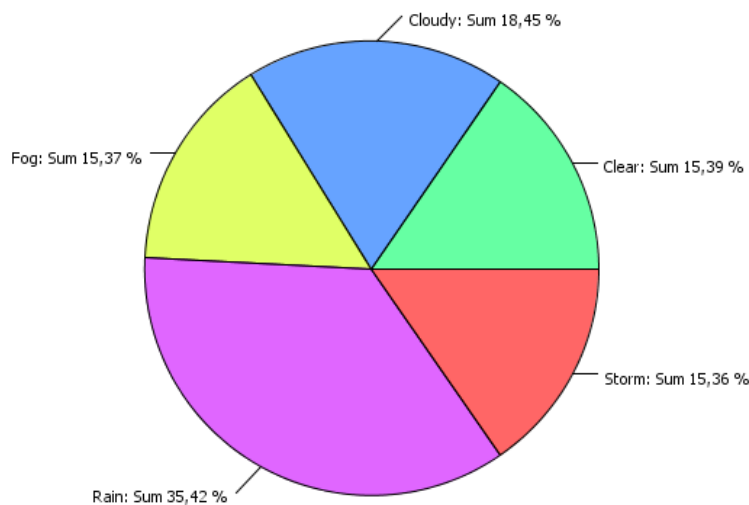
Na uvedeném obrázku byly modře označeny čtyři oblasti, které svou četností krádeží výrazně převyšují ostatní oblasti. Konkrétně se jedná o: nákupní centrum Géčko, těsné okolí hokejové arény, parkoviště u obchodního centra Nisa a okolí Harcovských kolejí. Na základě těchto výsledků byl v datech vytvořen v proudu modelu nový sloupec, který udává, ve kterém z kritických míst se čin odehrál.

Mimo nalezení kritických míst bylo dále žádoucí data prozkoumat a zjistit další informace, které nemusejí být na první pohled zřejmé. Tyto informace je možné získat pomocí dataminingových algoritmů, hledajících tzv. asociativní pravidla. Jedná se o nalazení jevů, které se v datech opakují.[3] Výsledky takovéto analýzy značně pomáhají vytvořit charakteristiku pachatele. V nástroji Knime je k dispozici uzel „association rule learner“, který používá algoritmus apriori. Tento uzel vyžaduje jako vstup vektor binárních proměnných. Podobně jako v úloze 5.8, bylo tedy potřeba z nominálních znaků, které tvoří vstup analýzy, vytvořit nejprve indikátorové proměnné. Pomocí kolekce indikátorových proměnných již bylo možné vytvořit potřebný vektor pro hledání asociativních pravidel. Pro analýzu byla nastavena minimální honfidence 70 % a dva prvky jako minimální délka velikosti vektoru. Výsledek analýzy vypadá takto:

Tabulka 5.3: Výsledky analýzy asociativních vazeb

Row ID	S Consequent	(...) Antecedent	D RuleConfidence%
Row0	OC_Nisa	[Fog]	71.4
Row1	odpoledne	[Fog]	71.4
Row2	odpoledne	[OC_Globus]	77.8
Row3	noc	[Harcov]	70
Row4	odpoledne	[OC_Nisa]	85.7
Row5	noc	[Ostatni]	84
Row6	noc	[Ostatni,Rain]	85.7

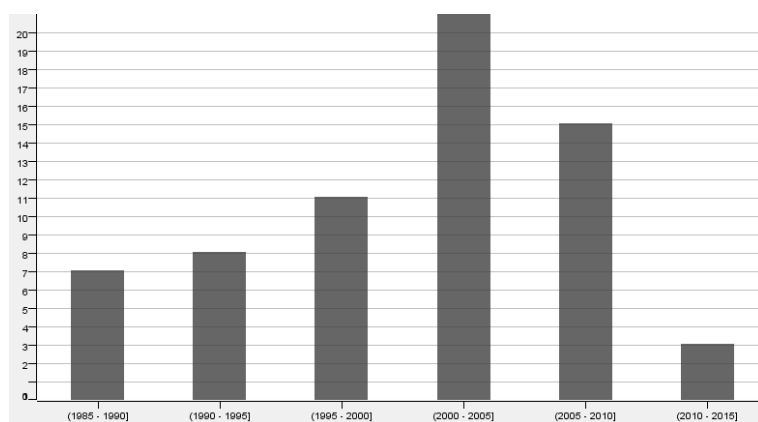
V tabulce je uvedeno sedm vazeb, které splňují výše uvedené podmínky. Oranžově byly označeny řádky, které mají zajímavý výsledek pro tuto úlohu. Z dat vyplývá, že zatímco u obou obchodních center dochází k většině krádeží v odpoledních hodinách, na ostatních místech k činu dochází zejména v noci. Důvodem může být například fakt, že zloděj v okolí obchodních domů může snadno kontrolovat pohyb vlastníka vozu. Zajímavá je také poslední uvedená vazba v tabulce. Zde bylo zjištěno, že více než 85 % krádeží, ke kterým dochází mimo obchodní centra, jsou provedeny přes noc a za deště. Na zobrazení zastoupení počasí v krádežích byl vytvořen výšečový graf:



Obrázek 5.18: Zastoupení počasí při krádeži

Lze vidět, že téměř 70 % krádeží se odehraje za nepříznivého počasí. Je tedy patrné, že toto počasí je pro zloděje vhodným spojencem. Důvodem může být snažší zamaskování hluků způsobených při samotné krádeži, či menší pohyb obyvatel města v ulicích.

Při vytvoření záznamu o krádeži se také ukládá značka a rok výroby vozidla. Tato data je možné prozkoumat pro zjištění základních statistických údajů. Histogram zobrazující krádeže podle roku výroby vozu vypadá následovně:



Obrázek 5.19: Histogram četnosti krádeží vozu dle roku výroby

V histogramu je možné sledovat určitý trend, kdy nejčastěji kradené vozy jsou zhruba deset až patnáct let staré. Starší a novější vozy mají zastoupení v histogramu výrazně nižší. Důvodem může být nižší poptávka v případě starších vozidel a vyšší náročnost odcizení v případě vozidel nových.

Zastoupení jednotlivých značek výrobců:

Tabulka 5.4: Zastoupení výrobců

Výrobce	Množství v %
Škoda auto	47,69
VW	15,38
Toyota	10,77
Fiat	10,77
Peugeot	9,72
Ford	6,15

Na základě výsledků této úlohy, bude vhodné zaměřit úsilí na kritické oblasti. V těchto místech by bylo vhodné například vytvořit nový kamerový systém. Podle závislosti na čase, kdy v lokalitách dochází ke krádežím je dále možné v tuto dobu lokalitu pravidelně sledovat hlídkou.

Závěr

V diplomové práci byla řešena problematika prevence kriminality s podporou dataminingových postupů v České republice s cílem přispět k v oblasti prevence. Spolupráce s Policií ČR začala účastí na mezinárodním workshopu „Mapy budoucnosti“ pořádaném Ministerstvem vnitra ČR na policejní akademii v Praze. Tam byla představena řada projektů prevence kriminality a jejich výsledky z EU a USA, které jsou inspirací pro analytiku a preventisty u nás. Z prezentovaných českých projektů vyplynulo, že spojení dataminingu a prevence kriminality, je v současné době v ČR ve stadiu vývoje a experimentů. Existuje několik lokálních řešení využívajících modely DM, například „Bezpečný Kolín“, ale obecně lze konstatovat, že nasazení dataminingu do práce kriminalistů analytiků není běžné. Provedli jsme rozsáhlou rešerši, či spíše studii o aktuálním stavu prevence kriminality s cílem porozumět používaným procesním postupům, používaným datovým zdrojům. Nalezli jsme několik bohatých zdrojů relevantních informací, které by mohly vykazat korelace mezi mírou kriminality a mohly by doplnit běžně používaná data při kriminálních analýzách. V praktické části jsme vyšli ze zkušeností, které jsme získali v citované rešeršní studii, na workshopu „Mapy budoucnosti“ a ve spolupráci s Městskou policií v Liberci a Policií ČR.

Vytvořil jsem nástroj, který by mohl sloužit v procesu prevence kriminality. Nástroj je komplexní tříúrovňový systém pro sběr dat v terénu, jejich následné databázové zpracování, až po nasazení dataminingových modelů založených na shlukové analýze a na vyhledávání asociačních pravidel popisujících chování pachatele. Pro práci kriminalistů a preventistů městské policie v terénu jsem vytvořil mobilní aplikaci pro operační systém android. Tato aplikace umožňuje sbírat data o vytipované trestné činnosti – krádeži aut, vloupání apod. a k těmto běžně sbíraným datům přidává automaticky další data, která dosud chybí nebo se musí obtížněji následně přidávat. Jde zejména o GPS souřadnice a informace o počasí. Pro získaná data byla navržena relační databáze systému MySQL, nad kterou pracuje webová aplikace sloužící jako dispečink a administrace. Třetí vrstvu jsem postavil v dataminingu

govém open-source nástroji Knime. Vytvořil jsem několik modelů, které demonstrují možné způsoby predikce kriminality a jejich zapojení do zvyšování bezpečnosti. Dalším výsledkem je ekonomický efekt vyplývající z efektivnější práce policejních složek, například optimálním nasazením policejních hlídek na kritická místa, které modely vytipovaly. Naše řešení umí také výsledky promítnout do map. Například vyznačí „hot spot“ místa, která slouží pro rychlou orientaci policie pro případný zásah či v prevenci vytipování míst pro budování kamerových systémů. Další vývoj vytvořeného nástroje by mohl směřovat k výraznějšímu přiblížení skutečným datům, to znamená ověření modelů na reálných datech. Ke splnění této ambice by bylo potřeba navázání oficiální spolupráce univerzity a místních složek městské policie a policie ČR. Díky rozsáhlosti dostupných veřejných dat, které jsme v této práci popsali, bylo by možné do systému zapojit další relevantní datové zdroje, poté modifikovat modely a zkoumat jejich vliv na výsledky modelování.

Literatura

- [1] MELOUN, Milan a Jiří MILITKÝ. Přednosti analýzy shluků ve vícerozměrné statistické analýze [online]. [cit. 2015-05-24]. Dostupné z: <http://meloun.upce.cz/docs/publication/152.pdf>.
- [2] KELBEL, Jan a David ŠILHÁN. Shluková analýza [online]. [cit. 2015-05-24]. Dostupné z: <http://www.fd.cvut.cz/personal/nagyivan/Projekty/Classification/ShlukovaAnalyza.pdf>
- [3] JIAWEI, Han, Jian PEI, Micheline KAMPER. Data Mining: Concepts and Techniques. Second Edition. Elsevier, 2011, 744 s. ISBN 978-0-12-381479-1
- [4] PLÍVA, Zdeněk; DRÁBKOVÁ, Jindra. Metodika zpracování diplomových, bakalářských a vědeckých prací na FM TUL. Vyd. 1. Liberec : Technická univerzita, 2007. 40 s. Dostupné z: http://www.fm.tul.cz/files/jak_psat_DP.pdf. ISBN 978-80-7372-189-3.
- [5] HRUŠKA, Lubor, Jiří FEJFARA a kol. Sborník příspěvků 2. Odborný workshop Mapy budoucnosti [online]. [cit. 2016-05-06]. Dostupné z: http://www.prevencekriminality.cz/evt_file.php?file=838
- [6] RUD, Olivia Parr. Datamining, Vyd.1. Praha: Computer Press, 2006, XVII, 329 s. ISBN 80-722-6577-6.
- [7] SKALSKÁ, Hana. Datamining a klasifikační modely, Vyd. 1. Hradec Králové: GAUDEAMUS, 2010, 154 s. ISBN 978-80-7435-088-7
- [8] BERKA Petr. Dobávání znalostí z databází. Praha: Academia, 2003, s.18. ISBN 80-200-1062-9
- [9] MAYER-SCHONBERGER, Viktor; CUKIER, Kenneth. Big Data. Vyd. 1. Praha: Computer Press, 2014, 256 s. ISBN 978-80-251-4119-9

Obsah příloženého CD

Příložené CD obsahuje zdrojové kódy k vytvořeným aplikacím, přílohy a diplomovou práci v elektronické podobě.