

Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky

Asociační analýza - příčiny a následky dopravních nehod

Michal Seiner

Bakalářská práce

2014

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: 2013/2014

ZADÁNÍ BAKALÁŘSKÉ PRÁCE (PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Michal Seiner**
Osobní číslo: **E090888**
Studijní program: **B6209 Systémové inženýrství a informatika**
Studijní obor: **Informatika ve veřejné správě**
Název tématu: **Asociační analýza příčina a následků dopravních nehod**
Zadávací katedra: **Ústav systémového inženýrství a informatiky**

Zásady pro vypracování:

Cíl práce: analýza příčin a následků dopravních nehod v okrese Pardubice a vyhledávání vzorů v podobě asociačních pravidel pomocí vhodného softwarového nástroje (Clementine, Modeler, Statistica, atd.). V BP budou teoreticky vymezeny základy asociační analýzy.

Náplň práce:

1. Základní popis dat
2. Charakteristika dopravních nehod
3. Metody asociační analýzy
4. Výsledky asociační analýzy

Rozsah grafických prací:

Rozsah pracovní zprávy:

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

Seznam odborné literatury:

- [1] BERKA, P. Dobývání znalostí z databází. Praha: Academia, 2003, 366 s. ISBN 80-200-1062-9
- [2] HAN, J., KAMBER, M. PEI, J. Data Mining: Concepts and Techniques. 3rd Edition, Academic Press 2011, 744 s. ISBN 978-01-2381-479-1
- [3] CHMELÍK, J. Dopravní nehody. Plzeň: Vydavatelství a nakladatelství Aleš Čeněk, 2009. 544 s. ISBN 978-80-7380-211-0

Vedoucí bakalářské práce:



Ing. Tomáš Kořínek

Ústav systémového inženýrství a informatiky

Datum zadání bakalářské práce: **30. září 2013**

Termín odevzdání bakalářské práce: **30. dubna 2014**



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

L.S.



prof. Ing. Jan Čapek, CSc.
vedoucí ústavu

V Pardubicích dne 30. září 2013

PROHLÁŠENÍ

Prohlašuji, že jsem tuto práci vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

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

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně Univerzity Pardubice.

V Pardubicích dne: 15. srpna 2014

Michal Seiner

PODĚKOVÁNÍ:

Tímto bych rád poděkoval svému vedoucímu práce panu Ing. Tomáši Kořínkovi za jeho odbornou pomoc, cenné rady a poskytnuté materiály, které mi pomohly při zpracování bakalářské práce. Dále chci poděkovat svým rodičům, kteří při mě vždy stáli a finančně mě podporovali. Neopomenu ani své přátele, jež mě podporovali a dodávali sílu k dokončení této práce.

ANOTACE

Tato bakalářská práce se zabývá analýzou příčin a následků dopravních nehod na území okresu Pardubice v letech 2006–2011. K řešení práce je použita asociační analýza konkrétně algoritmus Apriori. Vstupní data jsou zpracována pomocí tabulkového procesoru MS Excel a následně softwarovým nástrojem Clementine 12.0.

KLÍČOVÁ SLOVA

dopravní nehoda, asociační analýza, algoritmus Apriori, okres Pardubice

TITLE

Association analysis - causes and consequences of traffic accidents, Pardubice District

ANNOTATION

This bachelor thesis analyzes the causes and consequences of traffic accidents in the Pardubice district in 2006–2011. Association analysis (Apriori algorithm) is used for the solution of the above-mentioned problems. Input data is processed using MS Excel spreadsheet and software tool Clementine 12.0.

KEYWORDS

traffic accident, association analysis, Apriori algorithm, Pardubice District

OBSAH

1	SILNIČNÍ DOPRAVA A NEHODOVOST	11
1.1	DOPRAVNÍ NEHODA	12
1.2	PŘÍČINY DOPRAVNÍCH NEHOD	14
1.2.1	Objektivní příčiny dopravních nehod	14
1.2.2	Subjektivní příčiny dopravních nehod.....	15
1.2.3	Alkohol.....	16
1.2.4	Návykové látky.....	16
1.2.5	Léky.....	17
1.3	KLASIFIKACE DOPRAVNÍCH NEHOD.....	18
1.4	DĚLENÍ POZEMNÍCH KOMUNIKACÍ	19
1.5	BEZPEČNOST SILNIČNÍHO PROVOZU.....	20
1.5.1	Pravidla silničního provozu	20
1.5.2	Přímé a nepřímé ukazatele bezpečnosti silničního provozu	21
1.5.3	Národní strategie bezpečnosti silničního provozu 2011–2020.....	22
1.6	VÝVOJ NEHODOVOSTI	22
2	CHARAKTERISTIKA OKRESU PARDUBICE.....	24
2.1	OKRES PARDUBICE.....	24
2.2	SILNIČNÍ SÍŤ OKRESU PARDUBICE	25
3	DATA MINING	27
3.1	DATA MINING.....	27
3.2	METODOLOGIE.....	27
3.2.1	Metodologie SEMMA	27
3.2.2	Metodologie 5A.....	28
3.2.3	CRISP-DM	28
3.3	SPSS CLEMENTINE	30
4	ASOCIAČNÍ ANALÝZA	31
4.1	CHARAKTERISTIKY ASOCIAČNÍCH PRAVIDEL	31
4.1.1	Výhody a nevýhody asociačních pravidel	33
4.2	METODY ASOCIAČNÍ ANALÝZY	33
4.2.1	Apriori	33
4.2.2	CARMA	36
4.2.3	GRI	36
4.2.4	WEB	36
5	ASOCIAČNÍ ANALÝZA DOPRAVNÍ NEHODOVOSTI V OKRESU PARDUBICE	38
5.1	POROZUMĚNÍ PROBLÉMU.....	38
5.2	POROZUMĚNÍ DATŮM.....	38
5.2.1	Spojité atributy	39
5.2.2	Kategoriální atributy.....	39
5.3	PŘÍPRAVA DAT	42
5.4	MODELOVÁNÍ	44
5.4.1	Nehody bez osobních následků	44
5.4.1.1	Vedlejší analýzy	46
5.4.2	Nehody s osobními následky – lehká zranění	48
5.4.2.1	Vedlejší analýzy	49
5.4.3	Nehody s osobními následky – těžká zranění.....	51
5.4.3.1	Vedlejší analýzy	52
5.4.4	Nehody s osobními následky – usmrcení do 24 hodin	54
5.4.4.1	Vedlejší analýzy	55
5.4.5	Velikost hmotné škody	56
5.5	VYHODNOCENÍ VÝSLEDKŮ	59
	ZÁVĚR.....	61

SEZNAM TABULEK

Tabulka 1 – Limity pro drogy za volantem	17
Tabulka 2 – Kontingenční tabulka Ant a Suc	32
Tabulka 3 – Charakteristiky spojitých atributů	39

SEZNAM OBRÁZKŮ

Obrázek 1 – Silniční síť v ČR	12
Obrázek 2 – Fáze metodologie CRISP-DM	29
Obrázek 3 – Průběh algoritmu Apriori	35
Obrázek 4 – Pavučinový graf	37
Obrázek 5 – Datové typy jednotlivých atributů	43
Obrázek 6 – Nastavení uzlu Apriori	45
Obrázek 7 – Asociační pravidla s největší podporou (nehody bez osobních následků)	46
Obrázek 8 – Asociační pravidla s největší podporou (nehody s lehkým zraněním)	49
Obrázek 9 – Asociační pravidla s největší podporou (nehody s těžkým zraněním)	52
Obrázek 10 – Asociační pravidla s největší podporou (nehody s usmrcením do 24 hodin)	55
Obrázek 11 – Asociační pravidla s největší podporou (hmotné škody menší než 15 tisíc)	57
Obrázek 12 – Asociační pravidla s největší podporou (hmotné škody v rozmezí 15–45 tisíc)	58
Obrázek 13 – Asociační pravidla s největší podporou (hmotné škody od 45 tisíc a výše)	59

SEZNAM GRAFŮ

Graf 1 – Délka silniční sítě v okrese Pardubice	25
Graf 2 – Počet dopravních nehod za období 2006–2011 v okrese Pardubice	40
Graf 3 – Druhy dopravních nehod a jejich počet za období 2006–2011 v okrese Pardubice ..	40
Graf 4 – Zavinění dopravní nehody v okrese Pardubice v letech 2006-2011	41
Graf 5 – Počet nehod v okrese Pardubice v letech 2006-2011 v závislosti na příčině	41
Graf 6 – Místa nehod v okrese Pardubice v letech 2006-2011	42

ÚVOD

Silniční doprava zasahuje do našeho každodenního života, kdy většina z nás má už nějaké zkušenosti s řízením nebo výskytem dopravní nehody. Nemusí být přímo účastníkem dané události, ale slýchá historky, příběhy co se váží k dopravním nehodám. Tudíž nadávky, pomluvy, smích, smutek a překvapující situace ty k dopravní nehodě velmi pasují. Při debatách s kamarády a přáteli se rozebírají různé situace z cest, které občas vedou až na dlouhé teoretické diskuze co a jak by bylo lepší. I já sám mám už tyto zkušenosti za sebou.

Čím je silniční doprava déle na světě, tím se právě setkáváme s nárůstem dopravních nehod. Nejen u nás, ale i ve světě. Rok co rok jsou pravidla zpřísnována, silniční kontroly přibývají, narůstají ceny pokut, ale stále to není dost. Denně slýcháme, kde došlo k dopravním nehodám, kolik lidí bylo lehce či těžce zraněno, kolik lidí bylo při dopravních nehodách usmrceno, a jaká je celková hmotná škoda. Nepomáhá ani to, že jsou auta stále bezpečnější a vylepšují se všemožné vymoženosti na ochranu lidí. Problém nastává, že řidiči stále více riskují. Kupují si výkonnější vozy, které jak už bylo řečeno, mají více bezpečnostních opatření k záchraně životů. Podle mého názoru je uspěchanost a agresivita řidičů právě ten hlavní problém, proč dopravní nehody vznikají. U řízení by se mělo myslet na svoji bezpečnost, na bezpečnost spolujezdců, i na bezpečnost dalších účastníků silničního provozu. Ke zmírnění počtu nehod napomáhají i kampaně (např. "Nemyslíš, zaplatíš"), jejichž až někdy drastické, záběry ukazují možné důsledky dopravních nehod.

Cílem bakalářské práce je asociační analýza příčin a následků dopravních nehod v okrese Pardubice v letech 2006–2011. K potřebám práce je nejen důležité popsat a pochopit analyzovaná data, ale také charakterizovat související základní pojmy. Tato práce je rozdělena do pěti stěžejních kapitol.

První část se zabývá nehodovostí obecně, kdy je nejprve definován pojem dopravní nehoda. Dále jsou popsány možné příčiny vedoucí k dopravní nehodě a klasifikace dopravních nehod. V závěru kapitoly je popsána bezpečnost silničního provozu.

Druhá kapitola je věnována charakteristice analyzovaného území okresu Pardubice, jeho silniční síti a obecnému vývoji nehodovosti na území České republiky (ČR).

Třetí část se zabývá pojmem data mining (DM) a souvisejícím metodologiím (SEMMA, A, CRISP-DM).

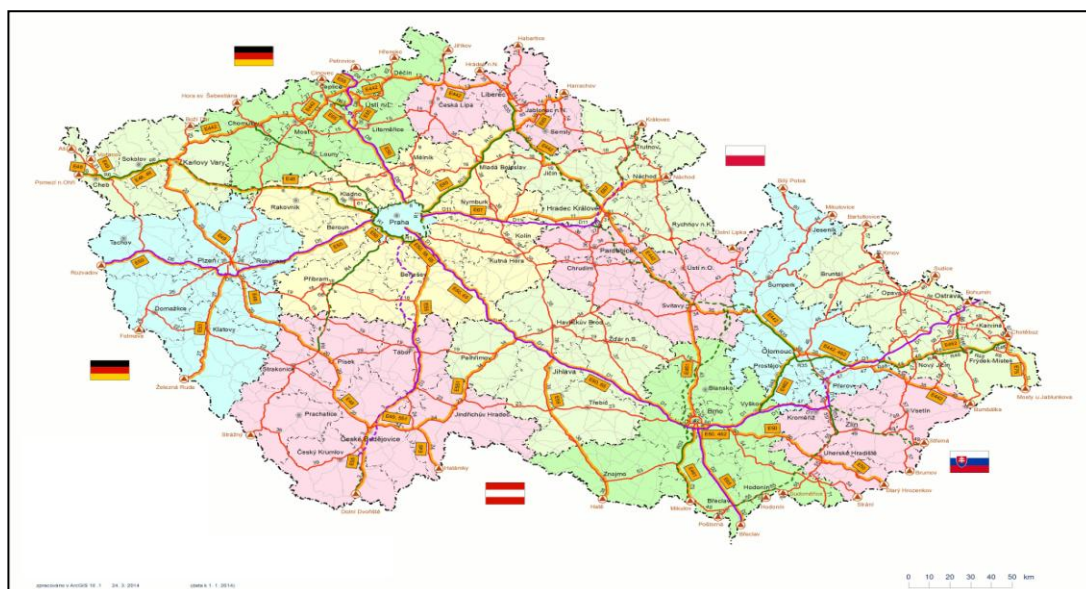
Předposlední část se věnuje asociační analýze. Jsou zde stručně charakterizována asociační pravidla včetně jejich výhod a nevýhod a též jsou popsány vybrané metody asociační analýzy.

Poslední část obsahuje asociační analýzu příčin a následku dopravních nehod v okrese Pardubice. K řešení zmíněné problematiky je využita metodika CRISP-DM. V té jsou nejprve definovány řešené problémy, následně jsou popsána zpracovávaná a analyzována data. Ty jsou v další fázi upravovány do konečné podoby. Závěrem jsou modelovány řešené problémy pomocí asociační analýzy a vyhodnoceny dosažené výsledky.

1 SILNIČNÍ DOPRAVA A NEHODOVOST

V této kapitole je definován pojem dopravní nehoda, jsou popsány jejich příčiny a možnosti klasifikace dopravních nehod. Dále je uvedena kategorizace silnic, stručně zmíněna problematika bezpečnosti silničního provozu a na závěr je popsán vývoj nehodovosti v ČR.

Historie silniční dopravy, resp. výstavby veřejné komunikační sítě, je stejně stará jako historie stavby měst, neboť ulice, náměstí a veřejné prostory sloužily odedávna k dopravě osob a nákladů. Silniční doprava využívající již dopravních prostředků jako jsou např. automobily, však patří k nejmladším a zároveň nejrychleji se rozvíjejícím odvětvím dopravy. Díky svým přednostem (operativnosti a rychlosti) úspěšně konkuruje některým starším odvětvím dopravy, především dopravě železniční. Ve světovém dopravním systému zajišťuje přepravu nákladů a osob převážně na krátké vzdálenosti a má rozhodující podíl na objemu světové nákladní i osobní přepravy. Ve většině ekonomicky vyspělých zemí již získala i vedoucí postavení ve výkonu nákladní přepravy. Uplatňuje se hlavně v přepravě vnitrostátní (vnitrooblastní), začíná se však prosazovat i v přepravě mezinárodní. Ve vyspělých zemích tvoří důležitý článek v systému tzv. „kombinované přepravy“, v němž navazuje na ostatní odvětví dopravy. V rozvojových zemích někdy představuje jediný moderní způsob pevninské dopravy, např. v některých afrických a asijských zemích, v nichž chybí železnice. Rozvoj silniční dopravy je provázen stálým růstem světové silniční sítě. Její délka se za posledních 70 let (rok 2014) více než zdvojnásobila, přičemž se změnila i její kvalita. Nepříznivý je však vliv klimatických podmínek, zejména na vlastní automobilový provoz. Silniční síť ČR je znázorněna na obrázku 1. Silniční doprava je nejméně bezpečným odvětvím dopravy a vyznačuje se nejvyšší nehodovostí. Ročně zemře na silnicích ve světě přes 400 000 lidí. [28]



Obrázek 1 – Silniční síť v ČR

Zdroj: [25]

1.1 Dopravní nehoda

Dle [9] je dopravní nehoda nepředvídaná, ale zpravidla předvídatelná událost, která vznikla během provozu na dopravní cestě a měla za následek škodu na životě, zdraví nebo majetku či jiný, zvláště závažný následek.

Tato definice, která je platná pro všechny druhy dopravních nehod, zahrnuje 3 základní pojmové znaky. Jsou to tyto tři:

- nepředvídatelnost, ale zpravidla předvídatelnost události,
- vztah dopravní nehody k dopravní cestě,
- následek v podobě způsobení škody na životě, zdraví nebo majetku nebo jiný zvláště závažný následek.

Jeden z těchto znaků, který se považuje za výchozí při každé dopravní nehodě, je považována skutečnost, že se jedná o událost nepředvídanou ale často předvídatelnou, která má určité meze, hranice. Toto předvídaní je možné z teoretického hlediska rozdělit na dvě základní roviny.

Předvídaní reálné

Jedná se o takové předvídaní konkrétné události, díky jejímuž průběhu s vysokou mírou pravděpodobnosti dojde v určitém prostoru a čase v případě, že existuje konkrétní příčina daná konkrétní událostí. Např. v silniční dopravě lze reálně předvídat, že při vysoké rychlosti vozidla

na zledovatělé silnici řidič při předjíždění v nepřehledné zatáčce na frekventovaném úseku vozidel způsobí dopravní nehodu. [9]

Předvídání abstraktní

Jedná se o předvídání abstraktně možné události, která může teoreticky nastat, pokud je naplněna řada souvisejících příčin a podmínek. Tento vznik událostí, kde pravděpodobnost faktického vzniku je velmi malá a je vázána na řadu příčin a podmínek.

Z pohledu teorie předvídání můžeme konstatovat, že neexistují události nepředvídatelné, ale jen události doposud nepředvídané člověkem.

Druhým znakem dopravní nehody je pohyb dopravního prostředku po dopravní cestě, tedy provoz. Jde o pohyb činitelů dopravy (řidiče nebo jiného účastníka dopravy a dopravního prostředku) po dopravní cestě v určitých objektivních podmínkách.

Třetím pojmovým znakem dopravní nehody je následek dopravní nehody. Způsobení škody na životě, zdraví, majetku nebo způsobení jiného, zvláště závažného následku. Za škodu je nutno považovat jednak reálnou újmu např. v souvislosti se vznikem obecného nebezpečí. [9]

Každá dopravní nehoda je charakteristická dvěma komponenty [9]:

a) nehodovým jednáním je pokládáno jednání účastníka dopravy, který svým konáním nebo opomenutím způsobil nehodovou událost

b) nehodou událostí je konkrétní projev dopravní nehody a mezi nejzákladnější patří [9]:

- Srážka – kdy jde o střet dvou či více účastníků silničního provozu, z nichž alespoň jeden se pohyboval za použití vozidla. Patří zde třeba čelní srážky, boční srážky, nárazy zezadu, náraz do pevné překážky, střet s chodcem.
- Havárie – dopravní nehoda se stává havárií, když je účastníkem pouze jedno vozidlo. Nebyla zapříčiněna žádná srážka s jiným vozidlem a neměl kontakt s žádnou pevnou překážkou. Jedná se například o smyk jedoucího vozidla, které skončí mimo silnici, převrácené na střeše.
- Jiné dopravní nehody – jde o nehody, které nemůžeme zařadit mezi havárie nebo srážky. Zde patří zranění či usmrcení osob uvnitř vozidla. Například zranění osob při prudkém brzdění.

Mezi nehodovým jednáním a nehodovou událostí existuje příčinná souvislost.

1.2 Příčiny dopravních nehod

Příčinou vzniku dopravní nehody je každý jev, bez něhož by následek v podobě dopravní nehody nenastal. Dopravní nehoda je výsledek působení činitelů člověka, vozidla a objektivních podmínek v provozu tj. dopravního provozu. Ve velké míře dochází ke kombinaci několika příčin a podmínek. Typické je, že příčina, která vede ke spáchání dopravního přestupku, existuje dříve, než řidič způsobí dopravní nehodu.

Níže je uvedeno 10 nejčastějších příčin, které vedly k dopravní nehodě na evropských silnicích v roce 2013 [27]:

- 1. místo:** celkem sedmnáct procent nehod zapříčiní překážka na vozovce nebo u vozovky.
- 2. místo:** 9,5 nehod ze 100 je způsobeno rychlou jízdou v zatáčkách.
- 3. místo:** sedm procent nehod je způsobeno špatným předjížděním.
- 4. místo:** méně než sedm procent případů má za následek blokování provozu.
- 5. místo:** pět a půl procenta má za následek nedání přednosti v jízdě.
- 6. místo:** pět procent má na svědomí jízda ve špatném jízdním pruhu, hlavně v levém.
- 7. místo:** pět procent nehod má na svědomí neopatrná jízda.
- 8. místo:** přibližně 4,5 řidičů ze 100 byli před nehodou rozptýleni.
- 9. místo:** čtyři procenta zapříčiní špatný stav vozovky.
- 10. místo:** tři procenta všech nehod zapříčiní smyk vozu.

Z předešlého textu si lze všimnout, že nejčastěji dopravní nehodu způsobí překážka na vozovce nebo v její blízkosti. Je zajímavé, že na předních pozicích žebříčku není žádné procento řidičů s alkoholem nebo návykovými látky obsaženými v krvi, o kterých v médiích slyšíme velmi často.

1.2.1 Objektívni příčiny dopravních nehod

Technický stav vozidla

Každé vozidlo chátrá, opotřebovává se materiál a vzniká koroze (hlavně na podvozku). Je proto důležité mít na paměti, že s vozidlem nejde jezdit po 10 letech stejně, jako když jsme ho kupovali. Hlavní body v zanedbání vlastníkem vozidla spočívá v závadě na brzdách, závadě v řízení či ve špatně udržovaných pneumatikách. Dnes už mají skoro všechny vozidla v sobě mnoho kontrol a čidel, které nám mají oznámit, zda s daným vozidlem pokračovat či ne.

Pokud budeme tyto signály ignorovat, vystavujeme nejen sebe, ale i široké okolí nebezpečím způsobené dopravní nehodou. Další důležitou objektivní příčinou silničních dopravních nehod, může být zvířectvo, ať už lesní nebo domácí. Nejčastěji dochází k nehodám v nočních hodinách a ne vždy se to obejde jenom s poškozením vozidla.

Dopravní prostředí

Pozemní komunikace může být další příčinou dopravních nehod. Kde její špatný stav, jako typickou příčinu lze uvést např. v zimním období špatná údržba pozemní komunikace nebo neoznačená překážka, má na svědomí už taky několik dopravních nehod. V této souvislosti jsou míněny všechny okolnosti bez přímého vlivu účastníka provozu na pozemních komunikacích, např. špatná viditelnost, či velmi hustý provoz. V ČR se často setkáváme s velkými reklamními tabulemi podél pozemní komunikace a právě jejich oznámení často strhává pozornost od řízení.

1.2.2 Subjektivní příčiny dopravních nehod

Selhání člověka při řízení dopravního prostředku může mít fatální následky. K nejnebezpečnějším však patří právní selhání člověka v podobě porušení konkrétních podmínek v provozu na pozemních komunikacích. V celém systému účastníků silničního provozu je nejproblémovější a nejvíce selhávajícím činitelem řidič dopravního prostředku. Je také potřeba zmínit řidičův momentální zdravotní stav, zdravotní potíže, psychické potíže, jeho věk a zkušenost s řízením. [9]

Další častou subjektivní příčinou dopravních nehod je tzv. „psychologická přednost“. Tento pojem znamená vzdání se přednosti v jízdě, která na křižovatce vyplývá řidiči z pravidel silničního provozu. Jev se uplatňuje zejména na křižovatkách neřízených světelnou signalizací, na nichž je přednost v jízdě vyznačena dopravními značkami. Řada dopravních nehod bývá způsobena vynucováním si přednosti tam, kde ji řidič přijíždějící do křižovatky po vedlejší silnici nemá. Způsob jeho chování, které je dáno především vysokou rychlostí vozidla při příjezdu ke křižovatce, má vyvolat u řidiče na hlavní silnici dojem, že druhý řidič nehodlá před křižovatkou zastavit, a že by tedy neměl z důvodu vlastní bezpečnosti na přednosti v jízdě trvat a raději se jí vzdát ve prospěch onoho agresivního jezdce. [20]

Speciální pozornost je nutná věnovat problematice „mikrospánku“. V mnoha případech je to oblíbeným argumentem řidičů, ovšem podle současných studií je mikrospánek stav několikavteřinové spánkové aktivity, který se vyskytuje u pacientů s narkolepsií. Zdravý člověk je vždy schopen si uvědomit příznaky blížícího se spánku. Pokud tedy řidič netrpí

chorobou, u které je mikrospánek jako výsledný projev, musí v dané věci odpovídat za následek, které v tomto stavu způsobí. [9]

1.2.3 Alkohol

Následky nehod zaviněných řidiči pod vlivem alkoholu, léků a jiných drog patří k těm nejtragičtějším. Současná právní úprava je ve vztahu ke konzumaci alkoholických nápojů během jízdy velice přísná, v ČR se uplatňuje nulová tolerance alkoholu, a to jak u řidiče motorového vozidla, tak i u cyklisty nebo jezdce na zvířeti. Za alkoholický nápoj se považuje lihovina, víno, pivo, a ostatní nápoje, které obsahují více než 0,5 objemového procenta alkoholu. [1]

V dnešní době jsou velmi oblíbené ochucené piva, u kterých není ani cítit nějaký ten alkohol, tak stále je to alkohol a za volant nepatří. Působení alkoholu začíná vždy až po nějaké době, a začíná to ve vývojově nejmladší části mozku. Proto jsou jako první ovlivněny zejména intelektuální činnosti člověka, logické myšlení, řeč apod. Vyvrcholením negativního působení alkoholu je narušení základních, pro život nezbytných funkcí nervové soustavy. Alkohol nám zpomaluje dobu při reakci a zhoršuje rozpoznávat ať už odhad vzdálenosti, rychlost jiných vozidel nebo úvahy řidiče co dělat při problémové situaci. Při konzumaci alkoholických nápojů se nám také zvyšuje pravděpodobnost zapříčinění dopravní nehody. Riziko každý zná a přece jen ho občas lidé zanedbávají a vystavuje riziku sebe i ostatní. To má za následek dopravní nehody kde je zranění, ztráty na životech nebo velmi vysoké majtkové škody, které ročně dosahují až miliard korun. [13]

1.2.4 Návykové látky

Nejen v poslední době je velmi propíraným tématem, nejen v médiích, problematika řízení motorových vozidel pod vlivem drog, a to především s ohledem na nebezpečnost takového počínání a na následky, které takto ovlivnění řidiči zaviní. Ze statistik nehodovosti vyplývá, že během roku 2013 bylo v ČR 213 dopravních nehod způsobeno pod vlivem omamných látek, na následky těchto nehod pak zemřelo 10 lidí a dalších 102 jich bylo zraněno.

Samotný pojem „droga“ český právní řád nezná, operuje totiž s pojmem návykové látky. Co se rozumí návykovou látkou, záleží na právní oblasti, v rámci které se pohybujeme. Jako návykové látky zákon označuje omamné látky a psychotropní látky. Řidiči pod vlivem účinku návykových a omamných látek reagují na situaci v provozu na pozemních komunikacích nepřiměřeným způsobem, mají např. motorické poruchy, mění rychlost jízdy, směr jízdy, mají často zhoršenou schopnost vnímat barvy signálních světel (zejména červenou). Přeceňují

vlastní schopnosti k jízdě, riskují a jsou bezohlední vůči ostatním účastníkům silničního provozu. Na obrázku 1 je vidět, jaké množství může mít řidič při řízení vozidla. Stanovené limity reflektují, od kdy již látka účinkuje již na zdravou populaci. Vedle toho ovšem záleží taktéž na váze řidiče a taky na tom, zda bere drogy pravidelně. [1]

Tabulka 1 – Limity pro drogy za volantem

Mezinárodní nechráněný název návykové látky v českém jazyce	Limitní hodnota návykové látky v krevním vzorku (ng/ml)
Konopí (9-THC)	2
Methamfetamin (pervitin)	25
Amfetamin	25
Extáze (MDMA)	25
3,4-Methylendioxyamfetamin (MDA)	25
Benzoylekgonin	25
Kokain	25
Morfin	10

Zdroj: [1]

1.2.5 Léky

Téma, o němž se příliš nemluví a přitom se týká všech lidí pohybujících se nejen po našich cestách. Zkuste si vzít své léky a odpovědět sami sobě zda víte, jaké léky byste neměli před jízdou užívat, jaká doba by měla uplynout mezi užitím léků a zasednutím za volant, případně jak poznáte rizikové léky? V dnešní velmi široké paletě léků není často jednoduché se zorientovat ani pro odborníka, natož pro laika. Jeho pozice je v oblasti léků ovlivňujících řízení složitější v tom, že již nejsou rizikové léky označované jasným symbolem (piktogramem) upozorňujícím na nebezpečí ovlivnění pozornosti. Pro někoho jsou navíc příbalové letáky psané často velmi malým písmem špatně čitelné. Schopnost bezpečně řídit motorová vozidla mohou ovlivnit také volně prodejné léky či léky na lékařský předpis. Pro řidiče jsou zvláště nebezpečné léky vyvolávající útlum nebo ospalost anebo naopak euforii, léky snižující sebekritičnost a sebekontrolu, ovlivňující látkovou rovnováhu v organismu a tím nervovou soustavu, krevní oběh, tlak nebo zrakové orgány. [21]

A teď si představte kombinaci drog, léků a alkoholu. V takovém stavu by řidič vůbec neměl pomyslet na řízení vozidla.

1.3 Klasifikace dopravních nehod

Doprava se dělí podle různých kritérií, jejichž podstatu obvykle tvoří prostředí, ve kterém se zrovna doprava uskutečňuje a způsob uskutečňování dopravy v daném prostředí. [9]

Podle prostředí, ve kterém leží dopravní cesta, dopravu dělíme na:

- pozemní,
- podzemní,
- vzdušná,
- vodní.

Podle způsobu uskutečňování dopravy na dopravní cestě v daném prostředí se doprava rozděluje na silniční:

- železniční,
- leteckou,
- plavební.

Podle vzdálenosti a místa provádění dopravy dopravu dělíme na:

- dálkovou,
- místní,
- městskou,
- speciální (např. závodní, školní apod.).

Podle druhu přepravy dopravu rozdělujeme na:

- nákladní,
- osobní

Z nadpisu této bakalářské práce je patrné zaměření se na pozemní, silniční a dálkovou dopravu. Proto je v následujícím textu charakterizována dopravní síť na území okresu Pardubice, taktéž jsou charakterizovány komunikace a okres Pardubice.

1.4 Dělení pozemních komunikací

Dle §2 zákona č.13/1997 jsou pozemní komunikace rozděleny následovně [29]:

- dálnice,
- silnice,
- místní komunikace,
- účelová komunikace.

O zařazení pozemních komunikací do příslušných kategorií a o změnách kategorie rozhoduje příslušný silniční správní úřad.

Dálnice

Pozemní komunikace určená pro rychlou dálkovou a mezistátní dopravu motorovými vozidly, která je budována bez úrovnových křížení, s oddělenými místy napojení pro vjezd a výjezd a která má směrově oddělené jízdní pásy. Je přístupná pouze silničním motorovým vozidlům, jejich nejvyšší povolená rychlost není nižší než 80 km/h (konstrukční rychlost vyšší jak 50 km/h). [11]

Silnice

Veřejně přístupná pozemní komunikace určená k užití silničními a jinými vozidly a chodci. Silnice tvoří silniční síť. Silnice se podle svého určení a dopravního významu rozdělují do těchto tříd. [11]

- Silnice I. třídy – určena zejména pro dálkovou a mezistátní dopravu.
- Silnice II. třídy – určena pro dopravu mezi okresy.
- Silnice III. třídy – určena k vzájemnému spojení obcí nebo jejich napojení na ostatní PK.

Místní komunikace

Je veřejně přístupná pozemní komunikace, která slouží převážně místní dopravě na území obce. Místní komunikace se dále rozdělují podle dopravního významu, určení a stavebně technického vybavení do čtyř tříd. [11]

- Místní komunikace skupiny A, kterou je zejména rychlostní místní komunikace s funkcí dopravní.

- Místní komunikace skupiny B, kterou je dopravně významná sběrná komunikace s omezením přímého připojení sousedních nemovitostí, funkce dopravně obslužní.
- Místní komunikace skupiny C, kterou je obslužná komunikace, s funkcí obslužnou.
- Místní komunikace skupiny D, kterou je komunikace se smíšeným provozem a komunikace s vyloučením motorového provozu.

Komunikace funkční skupiny D se dále dělí na podskupiny D1 a D2.

- D1: komunikace se smíšeným provozem
- D2: komunikace nepřístupné provozu silničních motorových vozidel

Rychlostní komunikace pro místní komunikace funkční skupiny A vychází svým příčným uspořádáním z dálnic a rychlostních komunikací, na které obvykle v intravilánu navazují. Rychlost je omezena na 80 km/h. [11]

Základní typy místních komunikací:

- MR – místní rychlostní.
- MS – místní sběrná.
- MO – místní obslužná.
- MOK – místní obslužná s krajnicí.

1.5 Bezpečnost silničního provozu

1.5.1 Pravidla silničního provozu

Pravidla silničního provozu jsou upravena zejména zákonem č. 361/2000 Sb. O provozu na pozemních komunikacích a o změnách některých zákonů (zákon o silničním provozu), ve znění pozdějších předpisů, v němž je též upravena oblast přestupků proti bezpečnosti a plynulosti provozu, bodové hodnocení řidičů nebo také oblast řídičských oprávnění.

K provedení zákona o silničním provozu bylo dále vydáno jedno nařízení vlády a několik vyhlášek v působnosti Ministerstva dopravy, Ministerstva vnitra a zdravotnictví. Konkrétně se jedná [22]:

- nařízení vlády č. 110/2001 Sb., kterým se stanoví další vozidla, která mohou být vybavena zvláštním zvukovým výstražným zařízením doplněným zvláštním výstražným světlem modré barvy,

- vyhlášku č. 30/2001 Sb., kterou se provádějí pravidla provozu na pozemních komunikacích a úprava a řízení provozu na pozemních komunikacích, ve znění pozdějších předpisů,
- vyhlášku č. 31/2001 Sb., o řidičských průkazech, ve znění pozdějších předpisů,
- vyhlášku č. 32/2001 Sb., o evidenci dopravních nehod,
- vyhlášku č. 277/2004 Sb., o stanovení zdravotní způsobilosti k řízení motorových vozidel, zdravotní způsobilosti k řízení motorových vozidel s podmínkou a náležitosti lékařského potvrzení osvědčujícího zdravotní důvody, pro něž se za jízdy nelze na sedadle motorového vozidla připoutat bezpečnostním pásem (vyhláška o zdravotní způsobilosti k řízení motorových vozidel), ve znění pozdějších předpisů,
- vyhlášku č. 218/2006 Sb., o úhradě za poskytnutí údajů z evidence dopravních nehod,
- vyhlášku č. 3/2007 Sb., o celostátním dopravním informačním systému,
- vyhlášku č. 124/2007 Sb., o vzoru paměťové karty řidiče,
- vyhlášku č. 82/2012 Sb., o provádění kontrol technického stavu vozidel a jízdních souprav v provozu na pozemních komunikacích (vyhláška o technických silničních kontrolách).

1.5.2 Přímé a nepřímé ukazatele bezpečnosti silničního provozu

Přímé ukazatele

Přímé ukazatele bezpečnosti silničního provozu (PUB) přímo reflektují bezpečnost silničního provozu na základě počtu a závažnosti následků dopravní nehody. Ukazateli jsou počet nehod, počet smrtelných, těžkých a lehkých zranění a velikosti hmotné škody. Každou silniční dopravní nehodu (SDN) můžeme ohodnotit z hlediska ekonomických ztrát. Vývoj nehodovosti v ČR v posledních letech značně závisel na opatřeních, která byla realizována s cílem zvýšení bezpečnosti silničního provozu. Každé nově zavedené opatření generuje určitý pozitivní efekt v podobě snížení počtu usmrcených.

Nepřímé ukazatele

Nepřímé ukazatele bezpečnosti silničního provozu (NUB) operují s okolnostmi či jevy, pomocí kterých lze posuzovat bezpečnost provozu na pozemních komunikacích a odhadovat

jejich další vývoj, aniž by byli nutné informace o SDN. Vycházejí z experimentálně ověřených vztahů mezi chováním a bezpečností provozu. V ČR jsou sledovány následující nepřímé ukazatele: rychlost vozidel, ochranné systémy, denní svícení vozidel, bezpečnostní odstupy a tak dále. [7]

1.5.3 Národní strategie bezpečnosti silničního provozu 2011–2020

Strategie byla schválena usnesením Vlády ČR ze dne 10. srpna 2011 a jedná se o národní strategii bezpečnosti silničního provozu na období 2011–2020. Je to samostatný materiál Ministerstva dopravy, který vytyčuje cíle, základní principy a návrhy konkrétních opatření směřujících k zásadnímu snížení nehodovosti na silnicích v ČR. Jako hlavní cíl je stanoveno snížení počtu usmrčených na úroveň evropských zemí a také oproti roku 2009 snížit počet těžce zraněných osob o 40%.

Strategie vytváří podmínky pro širší zapojení dalších resortů i všech ostatních subjektů, které mohou svou činností bezpečnost silničního provozu ovlivnit. K tomu, aby se nová strategie stala pro příští dekádu skutečně efektivním nástrojem pro zvýšení bezpečnosti silničního provozu je třeba účinným způsobem zapojit nejen veškeré zainteresované subjekty, ale i všechny další účastníky silničního provozu. [15]

1.6 Vývoj nehodovosti

Vývoj nehodovosti na silnicích v ČR prošel v posledních desetiletích několika významnými fázemi. Níže jsou uvedeny základní informace o vývoji BSP v období mezi lety 1970 až 2010. [20]

Mezi roky 1970 až 2010 se počet úmrtí následkem silničních nehod snížil o 60 % a počet zranění zhruba o 20%. Ve stejném období se počet vozidel více než zdvojnásobil. V letech 2000–2010 se počet úmrtí následkem silničních nehod snížil o 46 % a míra úmrtnosti (počet mrtvých následkem silničních nehod na milion obyvatel) se snížil o 40 %.

Vývoj dopravní nehodovosti lze ve sledovaném období rozdělit do 4 signifikantních fází:

V letech 1970–1986

Počet úmrtí následkem silničních nehod postupně snižoval a dosáhl své nejnižší úrovně v roce 1986. V té době byl počet smrtelných nehod na milion obyvatel v bývalém Československu srovnatelný s většinou vyspělých evropských zemí. Dle oficiálního hodnocení EHK OSN (Hospodářská komise OSN pro Evropu) bylo Československo jednou

z nejlepších zemí v souvislosti se snižováním počtu úmrtí na silnicích. Tento pozitivní vývoj byl vysvětlen následujícími faktory:

- zavedení prvních rychlostních limitů v extravilánu dnem 1. července 1979 (osobní automobily 90 km/h, těžká nákladní vozidla 70 km/h, motocykly 80 km/h);
- implementace nařízení č. 101/1981 Sb. Ministerstva vnitra, kterým je umožněno pozastavit řidičské průkazy řidičům, kteří nejsou schopni zaplatit pokutu za spáchaný dopravní přeštek. Toto nařízení, které nabylo účinnosti dne 1. ledna 1982, významně přispělo ke zlepšení bezpečnosti silničního provozu (BSP) v době své platnosti (toto nařízení už bylo zrušeno);
- koncem 60. let minulého století byl zahájen intenzivní rozvoj dálniční sítě (v roce 1980, dálnice byla otevřena mezi Prahou a Brnem);
- rychlejší rozvoj a modernizace vozového parku;
- silný tlak ze strany státu na prosazování aktivit pro zvýšení BSP.

V letech 1986–1996

Počty smrtelných nehod na silnicích se začaly mírně zvyšovat po roce 1986. Po roce 1989 se tento negativní trend ještě zvýšil a rok 1994 se stal nejhorším v počtu smrtelných silničních nehod. To lze vysvětlit především rychlým rozvojem motorizace a nesprávným pochopením "nově nabitě svobody" v souvislosti s politickým vývojem po 17. listopadu 1989.

V letech 1997–2003

Počet obětí silničních nehod se relativně ustálil a pohyboval se v určitém rozmezí. První výrazně pozitivní změnou bylo omezení rychlosti v intravilánu na 50 km/h (od 15. října 1997). Od 1. ledna 2001 bylo uzákoněno povinné denní svícení v zimním období a dále též uvedena v platnost přednost chodců na značených přechodech.

V letech 2004–2010

Po roce 2003 začal počet mrtvých následkem silničních nehod postupně klesat. V souvislosti s aktivním naplňováním stanovených cílů nově přijaté Národní strategie BSP byly posíleny pravomoci policie České republiky (PČR) byly iniciovány aktivity s cílem zlepšit silniční infrastrukturu na místní úrovni. Nejvíce pozitivních výsledků bylo dosaženo v roce 2006 (nejlepších od roku 1990) po zavedení bodového systému v červenci 2006. Přestože v následujícím roce 2007 nebylo dosaženo dobrých výsledků a nebylo tak navázáno na pozitivní trend vývoje nehodovosti, výsledky v letech 2008, 2009 a 2010 jsou opět povzbudivé.

2 CHARAKTERISTIKA OKRESU PARDUBICE

V následující kapitole je charakterizován okres Pardubice a jeho silniční síť.

2.1 Okres Pardubice

V roce 1960 okres Pardubice vznikl jako územní a správní celek sloučením převážné části území čtyř tehdy rušených okresů: Holic, Pardubice-město, Pardubice-okolí a Přelouč. Svoji rozlohou 880 km² je nejmenším okresem v kraji, na jeho území se podílí 19,5 %. Pardubický okres je rozdělen na 112 obcí, z toho 8 měst a jeden městys (Choltice). Hustota zalidnění činí 191 obyvatel na km² a je v kraji nejvyšší. V této oblasti pobývá 167 750 obyvatel a krajským městem po vzniku krajských částí v roce 2000 jsou Pardubice. [8]

Okres Pardubice se ve srovnání s jinými okresy ČR dlouhodobě vyznačuje relativně nízkou nezaměstnaností, přičemž nabízí pracovní příležitosti i zájemcům o práci ze sousedního okresu Chrudim. K rozvoji vzdělanosti také určitě přispívá velmi známá Univerzita Pardubice, která se řadí na 10. místo veřejných vysokých škol v ČR. Vznikla v roce 1994, má sedm fakult, šedesát sedm studijních programů a ve školním roce 2013/2014 počet studentů vzrostl už na hranici 10,5 tisíc. [26]

K nejvýznamnějšímu vodnímu toku celého okresu je řeka Labe s přítoky - řekou Chrudimkou a Loučnou. Pardubický okres patří k oblastem s ne příliš členitým povrchem a relativně nízkým výškovým rozpětím. Pevňá část území je součástí úrodné Polabské nížiny, proto takovou dominantou je hrad Kunětická hora, která se nachází nedaleko obce Brozany a Němčice s nadmořskou výškou 295 m nad mořem.

Z oblasti sportu má nadregionální význam dostihový závod Velká pardubická a motocyklové závody Zlatá přilba. Možnosti sportovního vyžití zejména pro cykloturistiku a pěší turistiku nabízí okolí bohdanečských rybníků a Kunětické hory [8]. Hlavní krajské město Pardubice se dokonce může hrdě hlásit k titulu "Hlavní město cyklistů", které získalo v roce 2014. [18]

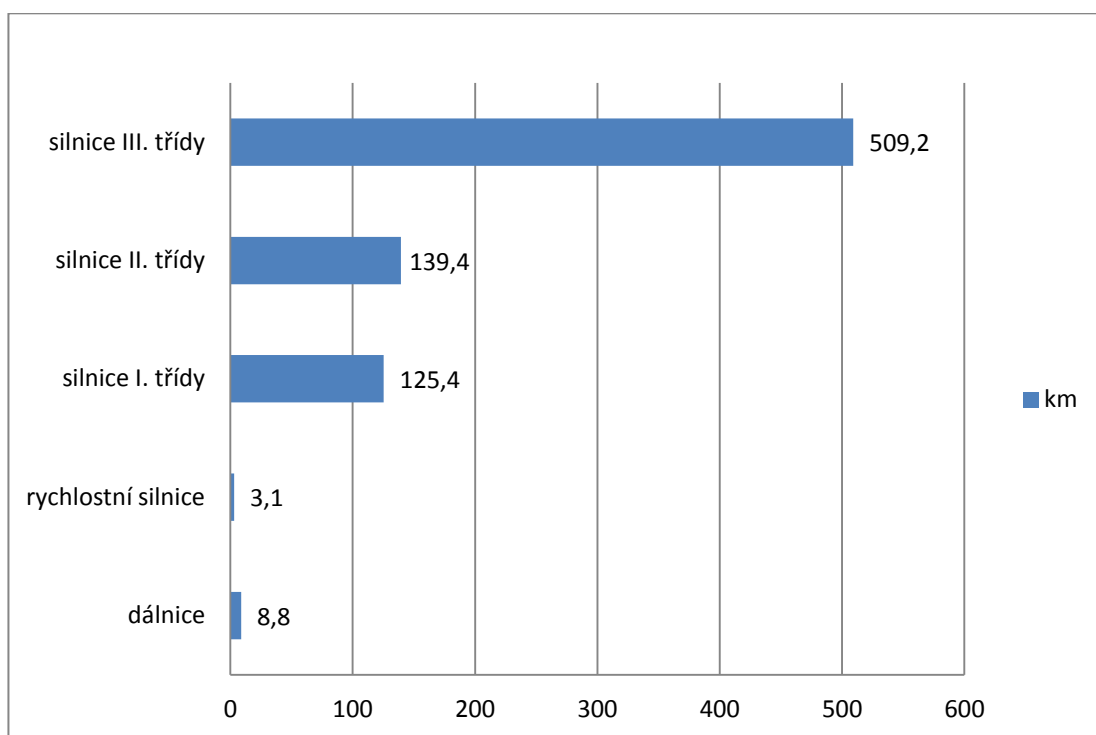
K turisticky atraktivním místům se můžou třeba pyšnit Pardubice historickým zámekem a zahradami. Nedaleko je možné navštívit pozdně gotický hrad na Kunětické hoře anebo přejet o kus dále a užívat si lázeňský areál v Lázních Bohdaneč. Na západě okresu Pardubice nalezneme areál hřebčína v Kladrubech nad Labem, který byl dokonce zvolen v 90. letech za národní kulturní památku. Pozoruhodné jsou také expozice afrického muzea v Holicích.

Na kulturní scéně Pardubického okresu určitě zaujímá Východočeské divadlo a Komorní filharmonie.

Pardubický okres má výhodnou polohu také z hlediska dopravního spojení. Jeho územím prochází ze západu na východ celostátně nejvýznamnější železniční trať Praha – Olomouc, která je modernizována výstavbou mezinárodního rychlostního koridoru. V železničním uzlu Pardubice ji křižuje trať Liberec – Havlíčkův Brod. Pardubice jsou již dostupné z hlavního města po dálnici D11 a navazující rychlostní komunikaci. Dále letiště v Pardubicích umožňuje smíšený vojenský a civilní provoz. [8]

2.2 Silniční síť okresu Pardubice

V pardubickém okrese je v roce 2013 silniční síť protkána 8 km dálnicí, která tu vznikla v roce 2006. Dále pak rychlostními silnicemi o délce 3,1 km, silnicemi I. třídy o délce 125,4 km, silnicemi II. třídy o délce 139,4 km, silnicemi III. třídy o délce 509,2 km a mnoha dalšími místními komunikacemi [23]. Graf 1 znázorňuje popsané skutečnosti.



Graf 1 – Délka silniční sítě v okrese Pardubice

Zdroj: [23]

Dálnice:

- D11 – je dlouhá pouhých 8 kilometrů a je na severní hranici okresu Pardubice. Její směr je Praha - Hradec Králové.

Silnice I. třídy:

- I/2 – směr Pardubice - Přelouč a dále pokračuje směr Kutná Hora,
- I/35 – směr Hradce Králové - Holice - Jaroslav a dále pokračuje směr Vysoké Mýto,
- I/36 – směr Nové město - Rohovládova Bělá - Pardubice - Sezemice - Holice a dále pokračuje směr Borohrádek,
- I/37 – směr Hradce Králové - Opatovice nad Labem - Pardubice a dále pokračuje směr Chrudim.

Silnice II. třídy:

- II/298 – Sezemice - Býšť - Bělečko,
- II/318 – Holice - Veliny - Rychnov nad Kněžnou,
- II/305 – Horní Jelení - Radhošť,
- II/322 – Kolín - Přelouč - Pardubice - Dašice - Vysoká u Holic,
- II/323 – Výrov - Pravy - Hradec Králové,
- II/324 – Hradec Králové - Opatovice nad Labem - Hrobice - Pardubice,
- II/333 – Hradec Králové - Lázně Bohdaneč - Přelouč,
- II/340 – Dašice - Úhřetická Lhota - Chrudim,
- II/342 – Valy - Svinčany - Chrudim,
- II/355 – Černá za Bory - Chrudim.

3 DATA MINING

V této kapitole je definován pojem data mining a stručně popsány vybrané metodologie.

3.1 Data mining

Data mining je pojem, pod kterým si lze představit velké množství technik sloužící k získávání znalostí z dat. Vzhledem ke stále rostoucí konkurenci v oblasti marketingu se DM stal nezbytnou činností vedoucí k udržení konkurenční schopnosti v každé fázi životního cyklu péče o zákazníka. Mezi obory pracující s technikami DM obecně patří pojišťovnictví, bankovníctví, veřejné služby, telekomunikace, cestovní ruch, zábava a jiné. V dnešní době nachází uplatnění i v oblastech veřejné správy, zdravotní péče, řízení lidských zdrojů a jiných. [19]

Data mining umožňuje pomocí speciálních algoritmů objevovat v datech strategické informace. Slouží manažerům k objevování nových skutečností, které pomáhají zaměřit jejich pozornost na podstatné faktory podnikání, slouží. Charakterizovat lze jako proces extrakce relevantních, předem neznámých nebo nedefinovaných informací z velmi rozsáhlých databází. Jedná se tedy o netriviální získávání implicitních, dříve neznámých a potenciálně užitečných informací z dat. [19]

3.2 Metodologie

Data miningový proces je závislý na vstupních zdrojích (zdrojové, softwarové, lidské, tedy finanční), které do něho vstupují. Možnost jak snížit finanční možnosti projektů, je provádět postupy standardizovanou formou. Z tohoto důvodu vznikly metodologie [2]

3.2.1 Metodologie SEMMA

Firma SAS vytvořila vlastní softwarový produkt na dobývání znalostí z databází Enterprise Miner, který vychází z jejich vlastní metodologie. Její název vznikl z počátečních písmen jednotlivých navazujících kroků [2]:

- sample – vybrání vhodných objektů, získání tzv. reprezentativního vzorku,
- explore – vizuální explorace a redukce dat,
- modify – úprava datové matice pomocí seskupování a transformace proměnných,
- model – využívání technik modelování pro získání požadovaných výsledků (neuronové sítě, regrese, shlukování, rozhodovací stromy...),

- assess – hodnocení využitelnosti a spolehlivosti vytvořených modelů.

3.2.2 Metodologie 5A

Tuto metodologii sestavila firma SPSS, název vznikl použitím akronymu jednotlivých kroků. Jedná se o následující kroky [2]:

- assess – posouzení potřeb projektu, jedná se o stanovení kontext, definování cílů a strategií,
- access – shromáždění potřebných dat a jejich příprava, získání a sestavení vhodných datových souborů,
- analyze – provedení analýz pomocí kterých jsou nalezeny odpovědi na otázky z prvního kroku,
- act – přeměna znalostí na akční znalosti, definování dodatečných otázek nebo doporučení a následné rozhodnutí,
- automate – převedení výsledků analýzy do praxe, aplikace rozhodnutí.

3.2.3 CRISP-DM

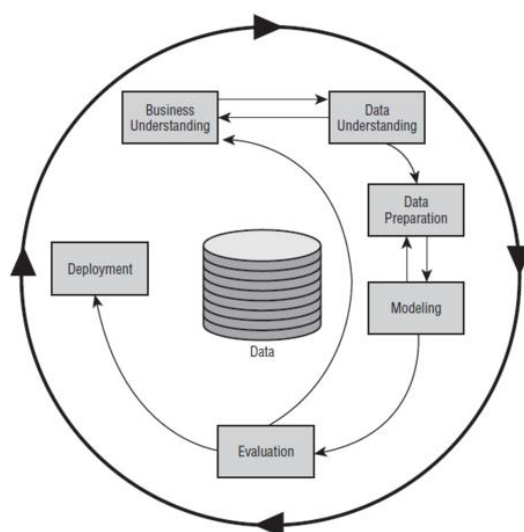
Nejznámější a nejpoužívanější je metodologie CRISP-DM, ta vznikla v rámci Evropského výzkumného projektu. Cílem bylo navrhnout univerzální postup (standardní model procesu získávání znalostí z databází), který by byl použitelný v nejrůznějších komerčních aplikacích. [19]

Životní cyklus každého projektu je podle této metodologie tvořen šesti fázemi. Jejich pořadí není pevně dané, dosažený výsledek z každé fáze ovlivňuje výstup fáze následující. Důležitou roli hraje fakt, že je možné se k jednotlivým fázím vracet. Metodologie CRISP-DM zobrazena na obrázku 2 je tvořena těmito následujícími fázemi [19]:

- **Porozumění problematice** (business understanding) - Je úvodní fáze zaměřená na pochopení cílů projektu a požadavků na řešení formulovaných z manažerského hlediska. Tato manažerská formulace musí být převedena do zadání úlohy pro dobývání znalostí z databází.
- **Porozumění datům** (data understanding) - Začíná prvotním sběrem dat. Následují činnosti, které umožní získat základní představu o datech, která jsou k dispozici (posouzení kvality dat, první „vhled“ do dat, vytipování zajímavých podmnožin záznamů v databázi...). Obvykle se zjišťují různé deskriptivní charakteristiky dat

(četnosti hodnot různých atributů, průměrné hodnoty, minima, maxima apod.), s výhodou se využívají i různé vizualizační techniky.

- **Příprava dat** (data preparation) - Zahrnuje činnosti, které vedou k vytvoření datového souboru, který bude zpracováván jednotlivými analytickými metodami. Tato data by tedy měla obsahovat údaje relevantní k dané úloze, a mít podobu, která je vyžadována vlastními analytickými algoritmy.
- **Modelování** (modeling) - Používá analytické metody s algoritmy pro dobývání znalostí. Obvykle existuje řada různých metod pro řešení dané úlohy, je tedy třeba vybrat ty nejvhodnější (doporučuje se použít více různých metod a jejich výsledky kombinovat) a vhodně nastavit jejich parametry. Jde tedy opět o iterativní činnost (opakovaná aplikace algoritmů s různými parametry), navíc, použití analytických algoritmů může vést k potřebě modifikovat data a tedy k návratu k datovým transformacím z předcházející fáze.
- **Vyhodnocení výsledků** (evaluation) - Se zabývá problémem, zda byly splněny cíle formulované na počátku projektu.
- **Využití výsledků** (deployment) - Vytvořením vhodného modelu celý projekt obecně nekončí. Dokonce i v případě, že řešenou úlohou byl „pouze“ popis dat, získané znalosti je třeba upravit do podoby použitelné pro podporu rozhodování. Podle typu úlohy tedy využití (nasazení) výsledků může na jedné straně znamenat prosté sepsání závěrečné zprávy, na straně druhé pak zavedení (hardwarové, softwarové, organizační) systému pro automatickou klasifikaci nových případů.



Obrázek 2 – Fáze metodologie CRISP-DM

Zdroj: [4]

3.3 SPSS Clementine

Pro řešení problematiky příčin a následků dopravních nehod v okrese Pardubice byl zvolen jako vhodný softwarový nástroj SPSS Clementine verze 12.0, který obsahuje celou řadu data miningových metod včetně takových, jež generují asociační pravidla.

Tento systém byl vyvinut britskou firmou Integral Solutions Ltd. v polovině 90. let. Počátkem roku 1999 firmu i její produkty přebírá výrobce statistického softwaru, firma SPSS. Tento program patří mezi přední komerční systémy pro dobývání znalostí, vychází z metodologie CRISP-DM. Systém nabízí mnoho metod pro deskriptivní i predikční úlohy jako jsou rozhodovací stromy, neuronové sítě, regrese či asociační pravidla. Vyznačuje se propracovaným způsobem ovládání, kdy uživatel pomocí „bloků“ pro zpracování, předzpracování, modelování nebo vizualizaci vytváří na pracovní ploše sekvenci řešení. [2]

4 ASOCIAČNÍ ANALÝZA

Kapitola se zabývá teoretickými východisky asociační analýzy a jejími metodami.

Smyslem asociační analýzy je hledání častých vzorů (vztahů), které jsou ukryté v datovém souboru. Tyto vzory jsou vyjadřovány pomocí asociačních pravidel, které popisují společně se vyskytující atributy. Asociační analýza a její obecné chápání je spjato s analýzou nákupního koše. Na základě této metody se zjišťují druhy zboží, které si uživatel pořizuje současně. Například soubor předmětů jako jsou mléko a chleba, které se objevují současně. Jde tedy o hledání vzájemných vazeb (asociací) mezi různými položkami sortimentu prodejny. [2], [9]

Asociační pravidlo je tvaru $A \Rightarrow B$, kde A, B jsou množiny položek $\{i_1, i_2, \dots, i_m\}$, vyskytující se v transakci T . Jejich. Asociační pravidla je zapisováno [9]:

$$i_1 \wedge i_2 \wedge \dots \wedge i_k \Rightarrow i_{k+1} \wedge i_{k+2} \wedge \dots \wedge i_{k+l}, \quad (1)$$

kde číslo $k+l$ označuje celkovou mohutnost asociačního pravidla K .

Ze vzorce (1) je patrné, že ke konstrukci asociačních pravidel se využívá syntaxe IF – THEN, která je vhodná k interpretaci získaných znalostí. [2]

4.1 Charakteristiky asociačních pravidel

Z datového souboru lze obecně generovat velké množství pravidel, to co je ovšem dělá zajímavými je, že jsou pochopitelné, použitelné, aktuální, ale hlavně platné s jistou mírou pravděpodobnosti. K tomu slouží právě metriky. [9]

U vytvořených pravidel je možné zjistit, kolik příkladů splňuje předpoklad a kolik závěr pravidla, kolik příkladů splňuje předpoklad i závěr současně, kolik příkladů splňuje předpoklad a nesplňuje závěr a obráceně. Tyto počty lze určit ze základního pravidla [2]:

$$\text{Ant} \Rightarrow \text{Suc}, \quad (2)$$

kde *Ant* (předpoklad, levá strana pravidla, antecedent) a *Suc* (závěr, pravá strana pravidla, sukcedent).

Pro n příkladů je sestavena následující tabulka 2.

Tabulka 2 – Kontingenční tabulka Ant a Suc

	Suc	¬Suc	Σ
Ant	a	b	r
¬Ant	c	d	s
Σ	k	l	n

Zdroj: [2]

Vysvětlivky k tabulce 1 jsou následující:

- $n(\text{Ant} \wedge \text{Suc}) = a$ je počet objektů splňujících současně předpoklad i závěr,
- $n(\text{Ant} \wedge \neg \text{Suc}) = b$ je počet objektů splňujících předpoklad a nesplňujících závěr,
- $n(\neg \text{Ant} \wedge \text{Suc}) = c$ je počet objektů nesplňujících předpoklad ale splňujících závěr,
- $n(\neg \text{Ant} \wedge \neg \text{Suc}) = d$ je počet objektů nesplňujících předpoklad ani závěr,
- $n(\text{Ant}) = a+b = r$, $n(\neg \text{Ant}) = c+d = s$, $n(\text{Suc}) = a+c = k$, $n(\neg \text{Suc}) = b+d = l$, $n = a+b+c+d$.

Na základě znalosti těchto hodnot (četnosti nebo také frekvenci kombinace) lze přistoupit k výpočtu různých charakteristik z pravidel a dosáhnout tak znalostí. Mezi základní charakteristiky patří podpora (support), spolehlivost (confidence), pokrytí (coverage) a kvalita (quality). [2]

Podpora

Podpora vyjadřuje počet objektů splňující předpoklad, tak i závěr. Tuto metriku je možné měřit jak v absolutní, tak i relativní podobě.

$$A \text{ resp } P(\text{Ant} \wedge \text{Suc}) = \frac{a}{a+b+c+d} \quad (3)$$

Spolehlivost

Jedná se o podmíněnou pravděpodobnost závěru, pokud platí předpoklad.

$$P(\text{Suc}|\text{Ant}) = \frac{a}{a+b} \quad (4)$$

Pokrytí

Vyjadřuje podmíněnou pravděpodobnost předpokladu, pokud platí závěr.

$$P(\text{Ant}|\text{Suc}) = \frac{a}{a+c} \quad (5)$$

Kvalita

Je vážený součet spolehlivosti a pokrytí.

$$\text{Kvalita} = w_1 \frac{a}{a+b} + w_2 \frac{a}{a+c} \quad (6)$$

Existují samozřejmě i další metriky jako kauzální podpora, kauzální spolehlivost, ujištění, zajímavost či závislost, avšak k potřebám práce postačí pouze základní metriky.

4.1.1 Výhody a nevýhody asociačních pravidel

Výhody asociačních pravidel jsou patrné již z předešlého textu, jednou z výhod jsou jejich srozumitelné a jasné výpočty popsanych charakteristik. Hledání pravidel není závislé na počtu atributů. Další nespornou výhodou jsou jasné a srozumitelné výsledky, které lze snad interpretovat. Asociační pravidla naleznou pravidla, která jsou skutečně obsažena ve zkoumaných datech. Na druhé straně velký počet atributů klade větší nároky na výpočet. Také s rostoucím počtem vstupních atributů roste možný generovaný počet kombinací, ten se redukuje nastavením kombinací do určité délky. Poslední dva zmíněné fakty vedou k větším nárokům na výpočet, a také na dobu zpracování. Kontraproduktivním je velký počet asociačních pravidel, především těch s nízkou hodnotou podpory. [2], [9]

4.2 Metody asociační analýzy

Základem všech metod (algoritmů) pro hledání asociačních pravidel je generování kombinací (konjunkcí) hodnot atributů. Během generování se prochází prostor všech přípustných konjunkcí. Jako metody jsou rozeznávány metoda do šířky, do hloubky a heuristická metoda. Pomocí první z metod generování do šířky se generují všechny kombinace délky jedna, poté délky dva a tak dále. Toto generování kombinací probíhá na základě délek. U generování do hloubky se vychází od první kombinace délky jedna, která se prodlouží o první kategorii dalšího atributu, do doby dokud je to možné. Oba předešlé způsoby mají tu nevýhodu, že neberou do úvahy analyzovaná data, proto jsou generovány kombinace nevyskytující se v datech. Poslední heuristická metoda generuje na základě četností, vytváří kombinace v pořadí dle jejich výskytu v dostupných datech. [2]

4.2.1 Apriori

Jednou z neznámějších a nepoužívanějších metod pro uplatnění asociačních analýzy a tedy hledání asociačních pravidel je metoda Apriori. Ta byla navržena R. Agrawalem v souvislosti s analýzou nákupního košíku, která se zabývá nalezením pravidel ve spotřebiteli nakupovaných produktech. Cílem tohoto algoritmu je hledání často se opakujících množin

položek (frequent itemsets). Jedná se kombinace (konjunkce) kategorií které dosahují předem zadané četnosti (podpory minsup) v datech. Při hledání kombinací délky k , které mají vysokou četnost, se využívají známé kombinace délky $k-1$. Při vytváření kombinace délky k , se spojují kombinace délky $k-1$. Jedná se o generování kombinací „do šířky“. Přitom pro vytvoření jedné kombinace délky k požadujeme, aby všechny její podkombinace délky $k-1$ splňovaly požadavek na četnost. [2]

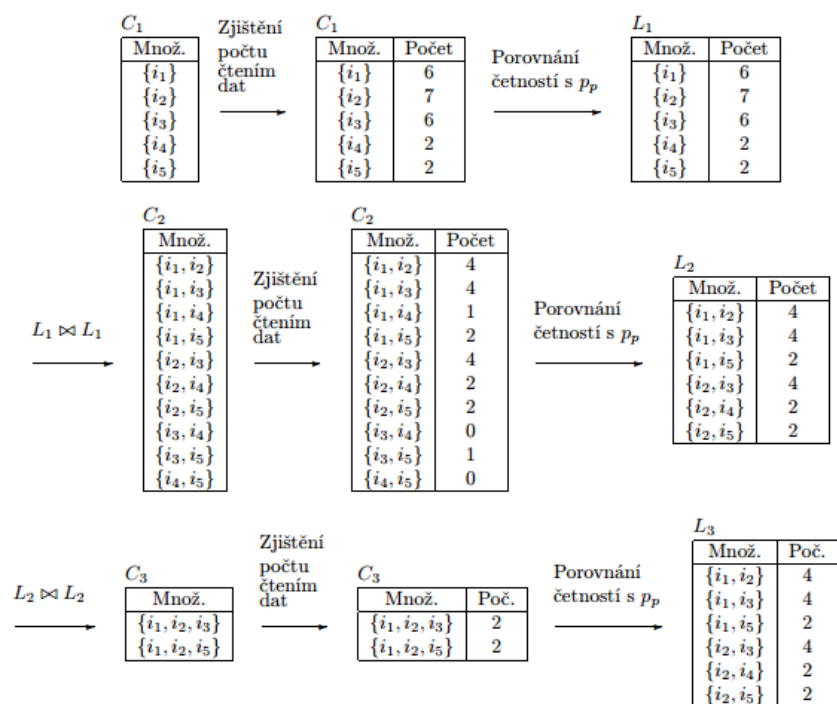
Algoritmus apriori probíhá na základě následujících kroků [2]:

1. do L_1 přiřaď všechny kategorie, které dosahují alespoň požadované četnosti
2. polož $k=2$
3. dokud $L_{k-1} \neq \emptyset$
 - 3.1. pomocí funkce apriori-gen vygeneruj na základě L_{k-1} množinu kandidátů C_k
 - 3.2. do L_k zařaď ty kombinace z C_k , které dosáhly alespoň požadovanou četnost
 - 3.3. zvětši počítadlo k

Kdy funkce apriori-gen(L_{k-1}) se skládá z následujících kroků:

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

Předcházející postup je na obrázku 3 zobrazen graficky.



Obrázek 3 – Průběh algoritmu Apriori

Zdroj: [3]

Po nalezení vyhovujících kombinací svou četností se vytvářejí asociační pravidla. Každá taková kombinace Com_b se rozdělí na všechny možné dvojice podkombinací Ant a Suc tak, že $Suc = Com_b - Ant$. Platí tedy, že Ant a Suc neobsahují stejnou kategorii ($Ant \cap Suc = \emptyset$) a zároveň $Ant \wedge Suc = Com_b$. Toto uvažované pravidlo $Ant \Rightarrow Suc$ má pak podporu, která je rovna četnosti kombinace Com_b . Spolehlivost pravidla se spočítá jako podíl četností kombinací Com_b a Ant , kdy četnost Ant je známa. [2]

Tento algoritmus je možné pomocí dalších postupů zefektivnit, a to následujícími metodami [3]:

- Hashování – použití této techniky může sloužit k redukci velikosti množiny kandidátských množiny položek C_k (pro $k > 1$).
- Redukce prohledávaných transakcí – vychází z toho, že pokud transakce (řádky tabulky) neobsahuje žádnou frekventovanou množiny velikosti k , nemohou obsahovat ani v $k+1$. Takové transakce mohou být označeny či odstraněny.
- Vzorkování – princip spočívá v náhodném výběru n transakcí z celé datové tabulky tak, aby se všechny vešly do paměti. Frekventované množiny se potom hledají na této množině transakcí.

4.2.2 CARMA

CARMA využívá asociační pravidla objevené algoritmem k nalezení asociačních pravidel v datech. Asociační pravidla jsou vytvářena ve známé formě IF – THEN.

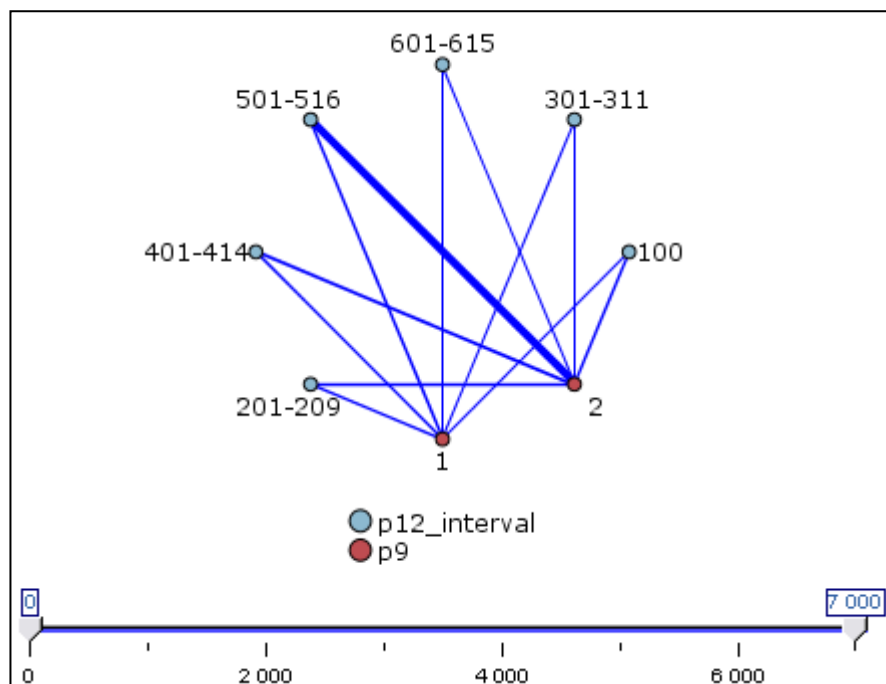
Model CARMA extrahuje soubor pravidel z dat, aniž by bylo nutné zadat vstupní nebo cílový atribut, tím se značně liší od algoritmu Apriori. To znamená, že vytvořená pravidla mohou být použita pro širší uplatnění. Například lze využít pravidla generované tímto algoritmem k nalezení seznamu výrobků nebo služeb (předchůdců), jejichž závěrem je produkt, který firma chce propagovat. Je možné vytvořit marketingovou kampaň na základě předchozích nákupů. CARMA je založena na „CARMA“ asociačních pravidlech algoritmu. Podpora je nastavena pro celé asociační pravidlo (podpora předchůdců i závěru), ne jako u Apriori, kdy je podpora určena pouze pro předchůdce. CARMA umožňuje vytvářet také pravidla s větším počtem závěrů. [9]

4.2.3 GRI

Metoda GRI (Generalized Rule Induction) slouží k objevování asociačních pravidel v datech, které jsou zaznamenávána ve formě IF – THEN. Pomocí této metody jsou vytvářeny pravidla z dat, které vykazují nejvyšší informační obsah. Informační obsah je měřen pomocí indexu, který bere v potaz podporu a spolehlivost vytvářených pravidel. Na rozdíl od metod Apriori a CARMA, které načítají tabulkové a transakční data, GRI vyžadují data v tabulkovém formátu. Asociační pravidla jsou obvykle poměrně snadně interpretovatelné na rozdíl o jiných metod. Pravidla se mohou překrývat, takže některé záznamy mohou vytvořit více pravidel. To umožňuje vytvořit sadu obecnějších pravidel než je tomu například u rozhodovacích stromů. Na rozdíl od předchozích metod dokáže pracovat s číselnými atributy jako předpoklady. [9]

4.2.4 WEB

Tato metoda je též nazývána jako pavučinový graf nebo síťový graf, kde atributy představují uzly a hrany mezi hodnotami dvou atributů představují sílu vztahu. [9] Příklad pavučinového grafu je zobrazen na obrázku 4, který vyjadřuje vztah mezi příčinami (modré body) a následky (červené body) dopravních nehod z analyzovaných dat.



Obrázek 4 – Pavučinový graf

Zdroj: [9]

5 ASOCIAČNÍ ANALÝZA DOPRAVNÍ NEHODOVOSTI V OKRESU PARDUBICE

Závěrečná kapitola se zaměřuje na modelování řešených problémů dopravní nehodovosti v okrese Pardubice, na níž je aplikována metodika CRISP-DM a její jednotlivé fáze, které byly popsány v předcházející kapitole.

5.1 Porozumění problému

Z názvu bakalářské práce i předcházející kapitoly je zřejmé, že k řešení problému bude použito asociační analýzy. Modelovány budou příčiny dopravních nehod, které vedou k následkům dopravních nehod. Jako následky jsou vybrány následující:

- bez osobních následků,
- s osobními následky,
 - lehce zranění,
 - těžce zranění,
 - usmrcení do 24 hodin po nehodě,
- velikost hmotné škody (finanční následky).

5.2 Porozumění datům

Zdrojová data k této práci o silničních nehodách v okrese Pardubice v letech 2006-2011 byla poskytnuta Ředitelstvím služby dopravní policie ČR. Každý z těchto záznamů byl zaznamenán zodpovědnou osobou do formuláře evidence nehod v silničním provozu a poté přepsán do elektronické podoby. Podoba formuláře pro zaznamenávání dopravních nehod a jednotlivé hodnoty atributů jsou uvedeny v příloze A.

Základní popis zpracovávaných údajů o dopravních nehodách je uveden v datovém slovníku, jež je součástí přílohy B.

Tato část práce je zaměřena na základní analýzu dat, pomocí které je získán větší přehled o zpracovávaných datech. Je zjištěno, jaká data jsou k dispozici, jejich rozsah, četnosti, minimální, maximální nebo průměrné hodnoty. Základní analýza dat byla provedena v tabulkovém procesoru MS Excel 2007.

Z vytvořeného datového slovníku (viz příloha B) lze vidět, že většina dat je kategoriálního typu (Nominal a Flag) a zbylé jsou spojitého typu (Range). Proto v první části budou řešeny spojitě atributy, především jejich minimální, maximální a průměrné hodnoty. Ve druhé části budou řešeny některé významnější atributy kategoriální, a to především počty jednotlivých kategorií a jejich poměry.

5.2.1 Spojité atributy

V tabulce 3 jsou zobrazeny základní charakteristiky spojitých atributů, ze kterých lze získat některé zajímavé skutečnosti. Například za sledované období bylo na území okresu Pardubice v důsledku nehody usmrceno 84 účastníků silničního provozu, 265 zraněno těžce a 2 448 zraněno lehce (atributy p13a, p13b a p13c). Co se týče škod, tak nejvyšší škoda na jednom vozidle dosáhla 2 milionů a průměrná 26 301 Kč a celková maximální hmotná škoda při jedné dopravní nehodě činí 4,74 milionu a průměrná 53 300 Kč (atributy p53¹ a p14). Největší havárie postihla 10 vozů (atribut p34).

Tabulka 3 – Charakteristiky spojitých atributů

	p13a	p13b	p13c	p14	p34	p53
Medián	0	0	0	300	2	100
Minimum	0	0	0	0	1	0
Maximum	2	4	11	47 400	10	20 000
Průměr aritmetický	0.01	0.02	0.20	533.39	1.79	263.01
Součet	84	265	2 448	6 470 099	24 405	3 190 312

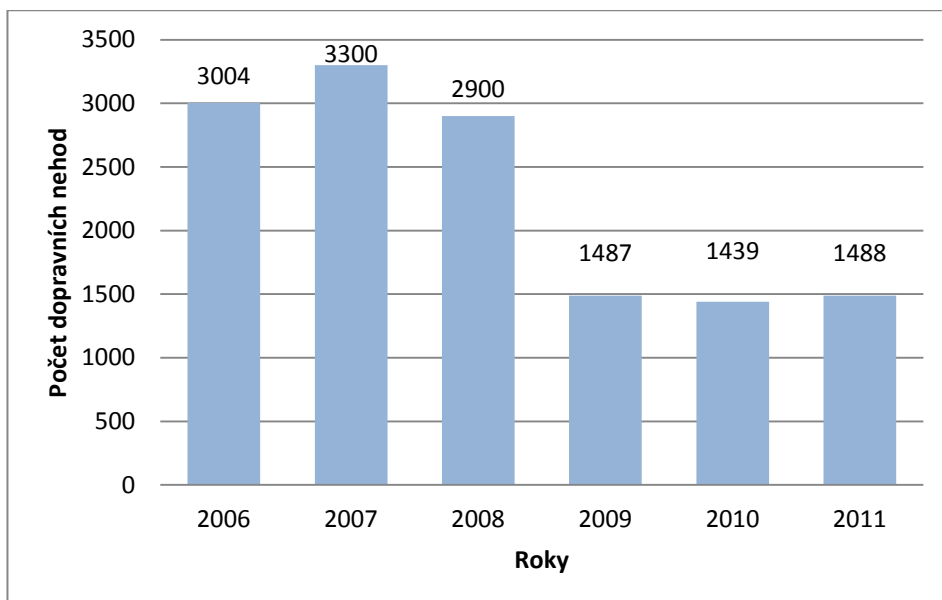
Zdroj: [24]

5.2.2 Kategoriální atributy

V následující podkapitole jsou analyzovány významnější kategoriální atributy, zbylé atributy včetně jejich počtů a procentuálních poměrů jsou přiloženy v příloze C.

Počet dopravních nehod v jednotlivých letech je zobrazen v grafu 2, kde je vidět výrazný propad mezi roky 2008 a 2009. To je způsobeno změnou zákona o povinnosti nahlášení dopravní nehody v důsledku výše škody z původních 50 tisíc na 100 tisíc Kč. V posledních letech po změně zákona je vidět poměrně konstantní počet dopravních nehod.

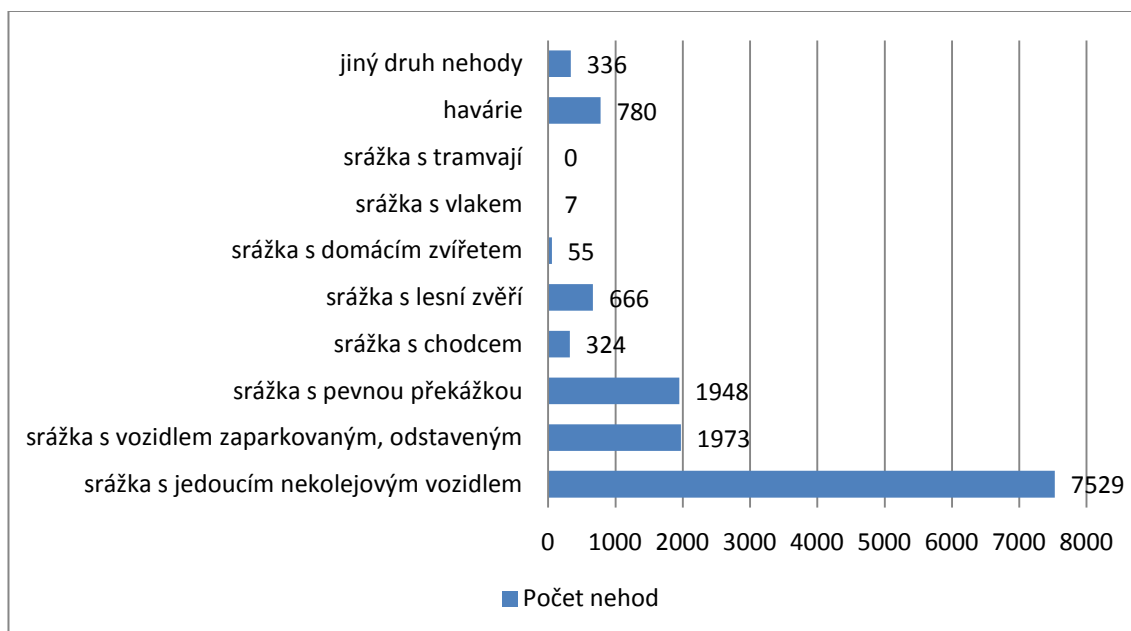
¹ Atributy p14 a p53 jsou zaznamenávány v 100 Kč.



Graf 2 – Počet dopravních nehod za období 2006–2011 v okrese Pardubice

Zdroj: [24]

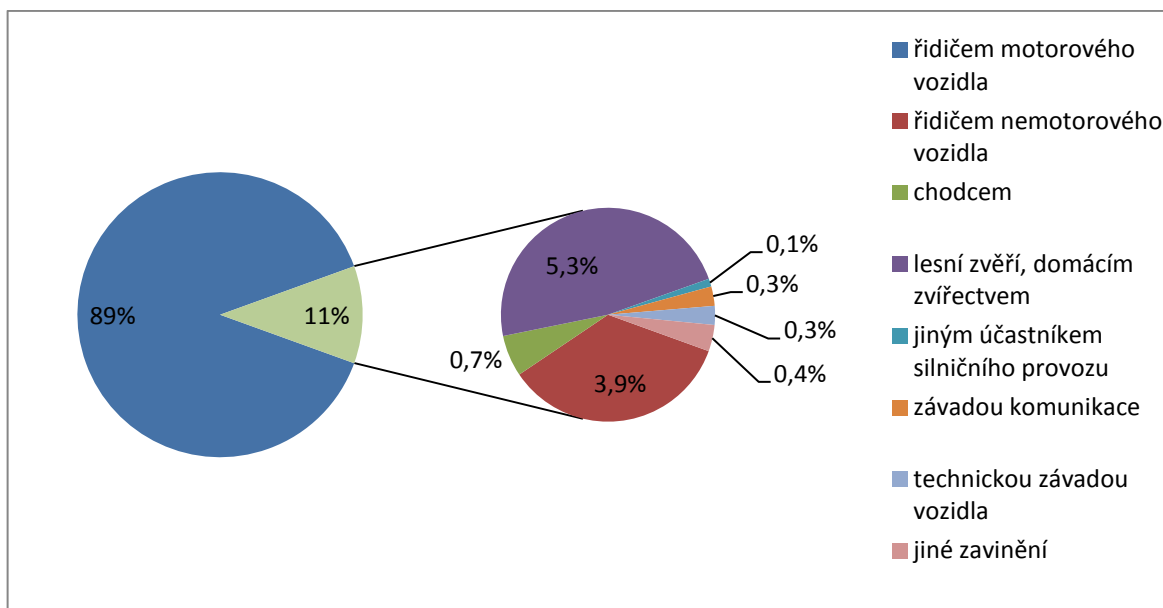
Na následujícím grafu 3 je charakteristika dopravních nehod v závislosti na druhu nehody. Nejčastěji je vznik nehody v důsledku střetu s jiným vozidlem, a to v 7 529 případech (55,28%). Dalšími významnými příčinami jsou srážka se zaparkovaným vozidlem 1 973 (14,48%) případů a srážka s pevnou překážkou 1 948 případů (14,3%).



Graf 3 – Druhy dopravních nehod a jejich počet za období 2006–2011 v okrese Pardubice

Zdroj: [24]

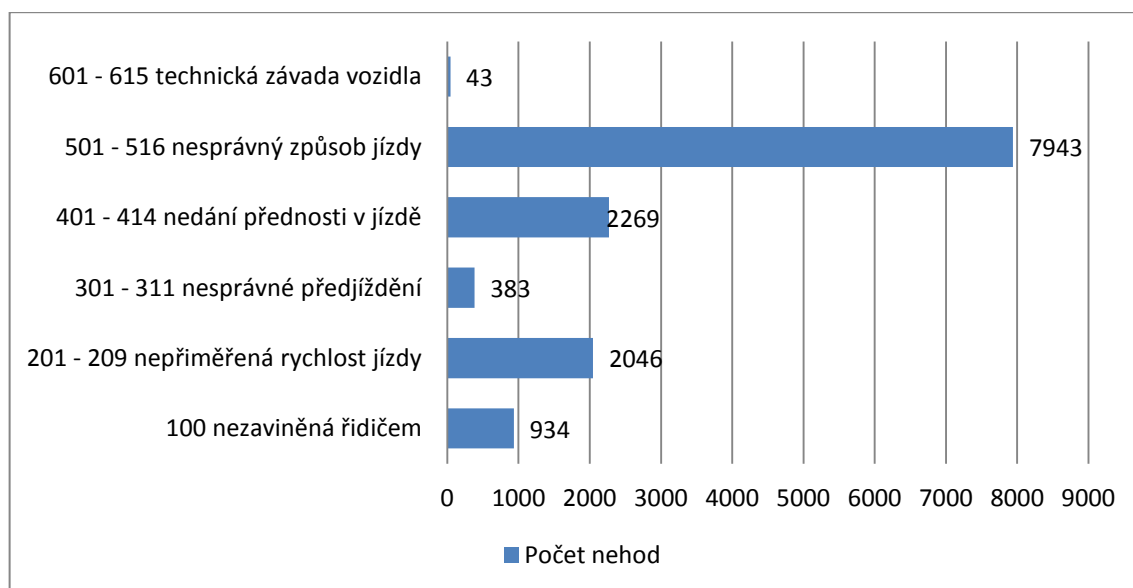
Graf 4 znázorňuje důvod vzniku dopravní nehody. Převážnou většinu, a to 89% (12 114) všech nehod je způsobeno řidičem motorového vozidla. Mezi ostatní nejpočetnější důvody patří především řidič nemotorového vozidla (cyklista nebo jezdeck na koni) nebo zvíře.



Graf 4 – Zavinění dopravní nehody v okrese Pardubice v letech 2006-2011

Zdroj: [24]

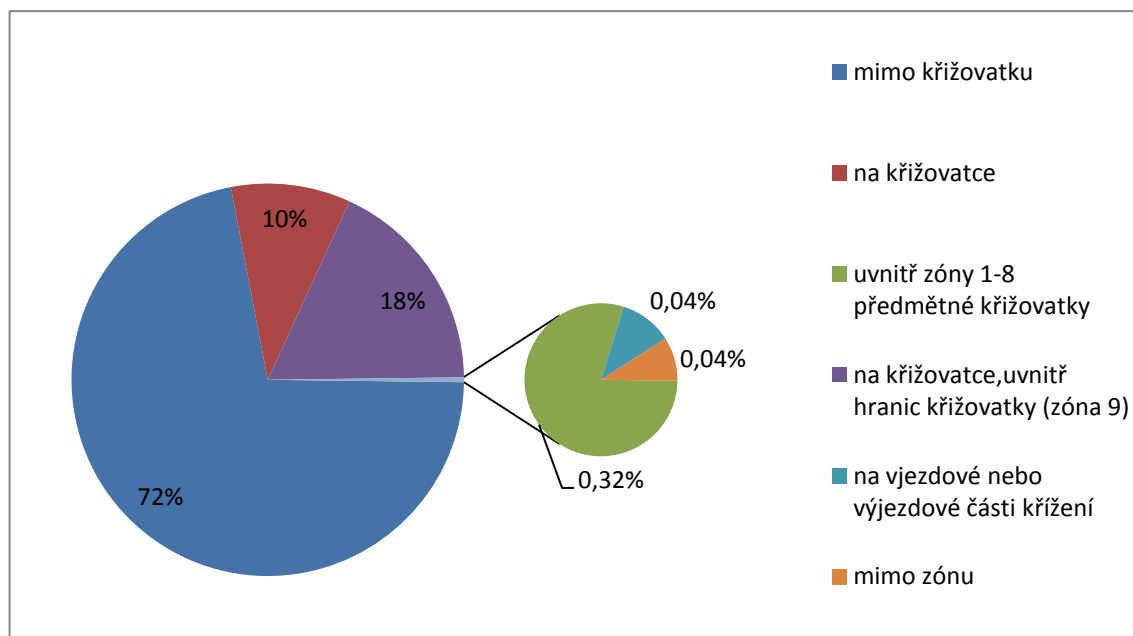
Graf 5 zobrazuje počty dopravních nehod podle jejich příčiny. Nejčastější příčinou je nesprávný způsob jízdy, který vedl k 7 943 (58,32%) všech nehod. Dalšími nejpočetnějšími příčinami jsou nedání přednosti v jízdě 2 269 případů (16,66%) a nepřiměřená rychlost jízdy ve 2 046 případech (15,02%).



Graf 5 – Počet nehod v okrese Pardubice v letech 2006-2011 v závislosti na příčině

Zdroj: [24]

V grafu 6 jsou znázorněny počty a poměry míst, na kterých se nehody v okrese Pardubice ve sledovaném období udály. Největší počet nehod vznikl mimo křižovatku 9 776 (72%), následně na křižovatce (zóna 9) 2 446 (18%) a na křižovatce 1 342 (10%). Ostatní možnosti jsou v tomto případě zanedbatelné.



Graf 6 – Místa nehod v okrese Pardubice v letech 2006-2011

Zdroj: [24]

5.3 Příprava dat

Zdrojová data jsou rozdělena do 6 listů, kdy každý vyjadřuje dopravní nehody spadající do určitého roku. Výsledná datová matice vznikla sloučením jednotlivých listů ze souboru data_nehody.xls. Jedná se celkově o 13 618 dopravních nehod se 48 atributy, kde každý řádek vyjadřuje jednu dopravní nehodu na území okresu Pardubice mezi roky 2006 a 2011.

Co se týče chybějících hodnot, ty opravovány nebyly z důvodu nedostupnosti příslušných podkladů pro doplnění těchto hodnot. Žádné nové atributy vytvářeny nebyly. Opravovány byly pouze datové formáty u některých atributů. Například atribut p2a (datum uskutečnění nehody) nebyl ve všech případech nastaven jako datový typ datum. Zbylé atributu jsou nastaveny na datový typ obecný, což je v některých případech nežádoucí. Například u atributu p47 (rok výroby vozidla), který by měl být zaznamenáván jako dvojciferné číslo, chybí u vozidel roku výroby 2000–2011 znak nuly na počátku. V tomto případě je vytvořen formát, který nulu zobrazuje. Stejný problém je opraven u atributů p17, p27, p35, p44, p45a, p47, p48a a p52, které musí obsahovat dvojciferné hodnoty. V poslední řadě byl upraven atribut p38, který vyjadřuje, na kolikátém kilometru komunikace se stala dopravní nehoda.

Zaznamenávané údaje jsou v desítkách metrů pro lepší přehlednost jsou převedeny na kilometry.

Vytvořená datová matice z předcházejícího odstavce je uložena ve formátu xls. Následně je převedena do formátu csv. oddělený středníkem, který bude vstupem do programu Clementine. Důležitou úpravou je nahrazení čárek v atributu p38 desetinnými tečkami, protože software není kompatibilní s čárkami jakožto oddělovačem desetinných míst. Výsledný soubor je uložen jako data_nehody.csv.

Nyní se činnosti přesouvají do prostředí SPSS Clementine. Pomocí uzlu Varfile na záložce File je vytvořený soubor data_nehody.csv nahrán do prostředí Clementine. Zde je nastaven jako oddělovač možnost „Other“ a doplněn středník. Na následující kartě Data jsou formáty jednotlivých atributů tak, jak je software rozlišil. Většina atributů je chápána jako integer (číslo), ale v případech, kdy hodnoty atributů začínají znakem 0, jsou programem tyto počáteční znaky vynechány. Proto jsou u atributů p17, p27, p35, p44, p45a, p47, p48a a p52 změněny datové typy na string (text) Změněn je také atribut p2a na formát datumu. Na další záložce Filter je možné odfiltrovat nepotřebné atributy. Protože zatím není známo, které atributy budou potřebné, jsou vyfiltrovány pouze identifikační údaje, jmenovitě atributy X, Y a p1. Předposlední ze záložek slouží k nastavení datových typů v rámci projektu, což je zobrazeno na obrázku 5.

Field	Type	Field	Type
p36	Set	p20	Set
p37	Set	p21	Set
p38	Range	p22	Set
p2a	Set	p23	Set
den	Set	p24	Set
čas	Set	p27	Set
p6	Set	p28	Set
p7	Set	p34	Range
p8	Set	p35	Set
p9	Set	p39	Set
p10	Set	p44	Set
p11	Set	p45a	Set
p12	Set	p47	Set
p13a	Range	p48a	Set
p13b	Range	p49	Flag
p13c	Range	p50a	Set
p14	Range	p50b	Set
p15	Set	p51	Set
p16	Set	p52	Set
p17	Set	p53	Range
p18	Set	p55a	Set
p19	Set	p57	Set
		p58	Set

Obrázek 5 – Datové typy jednotlivých atributů

Zdroj: [vlastní]

Upraven je však atribut p12 (hlavní příčiny nehody), jehož hodnoty jsou v datovém slovníku popsány intervalem, zatímco v datech nejsou. Úprava je provedena pomocí uzlu Reclassify, kde příslušným hodnotám jsou nastaveny hodnoty odpovídající datovému slovníku.

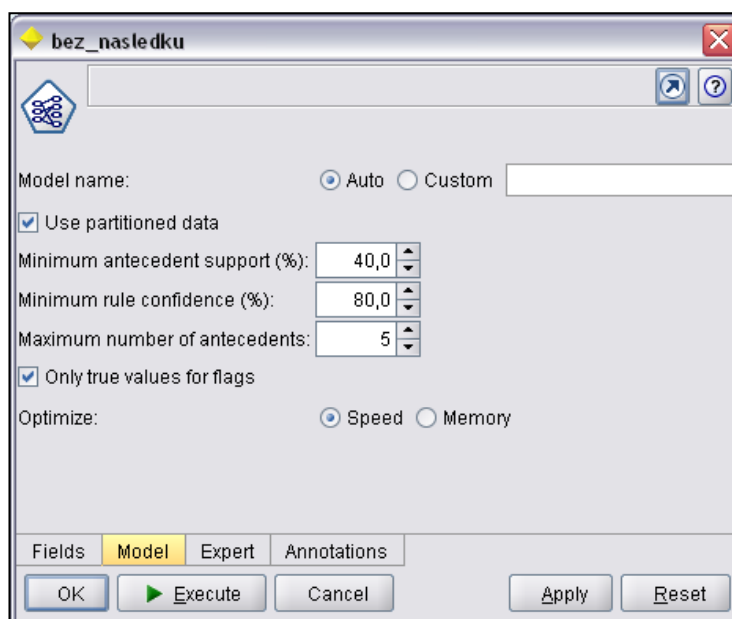
Takto připravená data jsou vstupem do následující fáze modelování.

5.4 Modelování

Následující kapitola je zaměřena na modelování definovaných problémů z kapitoly porozumění problémů pomocí asociační analýzy na vytvořených datech z předešlé kapitoly. Nejprve je analyzován problém definovaný v kapitole 5.1 (nehody bez osobních následků, s osobními následky a velikost finanční škody) v rámci těchto hlavních problémů jsou dále analyzovány problémy vedlejší. Například pro nehody bez osobních následků jsou jako vedlejší problémy analyzovány nehody, které se staly o víkend, které se staly v důsledku srážky se zvěří nebo například, které byly způsobeny přítomností alkoholu u řidiče. Tyto vedlejší analýzy slouží k získání dalších informací o daných následcích.

5.4.1 Nehody bez osobních následků

První z řešených problémů se zabývá odhalením faktorů, které vedou k dopravní nehodě, při níž nedošlo k osobním následkům na zdraví účastníků silničního provozu. Závěrem (Consequent) je nehoda bez osobních následků. Tyto údaje se získají z atributu p9 = 2, pomocí uzlu Derive je vytvořen nový atribut bez_nasledku, kde hodnotě p9 = 2 je nastavena pravda a ostatním hodnotám nepravda. Následně jsou pomocí uzlu Select vybrány pouze záznamy splňující pravdu. Dalším krokem je výběr předpokladů (Antecedents), které mají vliv na vznik dopravní nehody. Obecně by se daly předpoklady rozdělit do dvou skupin na vnitřní (na straně řidiče) a vnější (podmínky při řízení), nicméně budou analyzovány dohromady. Jako předpoklady byly zvoleny atribut p6 (druh dopravní nehody), p10 (zavinění nehody), p11 (alkohol u řidiče), p12_interval (hlavní příčiny nehody), p16 (stav vozovky v době nehody), p18 (povětrnostní podmínky), p19 (viditelnost), p57 (stav řidiče) a p58 (vnější ovlivnění řidiče). Zbylé atributy nebudou uvažovány, proto jsou pomocí uzlu Filter odfiltrovány. Připojen je uzel Type, ve kterém jsou nastaveny vstupní atributy hodnotou „In“ a výstupní atribut hodnotou „Out“. Závěrem je připojen uzel Apriori, ve kterém jsou nastaveny předpoklady a závěr. Na obrázku 6 je nastavení uzlu Apriori, kdy se nastavují minimální hodnoty podpory a spolehlivost, kterých musí model dosáhnout, aby vypsal asociační pravidla. Dále je zde možnost nastavení maximálního počtu předpokladů.



Obrázek 6 – Nastavení uzlu Apriori

Zdroj: [vlastní]

Za sledované období se uskutečnilo 10 791 dopravní nehod, kdy následkem nebyla újma na zdraví. Nastavením uzlu Apriori, viz předchozí obrázek, vzniklo 215 asociačních pravidel. Na obrázku 7 jsou zobrazeny asociační pravidla s nejvyšší podporou, tabulka v záhlaví obsahuje Consequent (závěr), Antecedent (předpoklad), Support % (počet záznamů splňující předpoklad i závěr, viz kapitola 4.1), Instances (shodné se support, zaznamenávané však v absolutních hodnotách), Confidence (podmíněná pravděpodobnost závěru pokud platí předpoklad, viz kapitola 4.1), Rule Support (podíl záznamů, pro které je celé pravidlo, předpoklad i závěr pravdivé), Lift (poměr podpory pravidla k podmíněné pravděpodobnosti závěru) a Deployability (procento trénovacích dat splňující předchůdce a nesplňující závěr). Z vytvořených pravidel je patrné, že nejčastěji dochází k dopravní nehodě bez osobních následků pokud $p_{10} = 1$ (nehoda způsobena řidičem motorového vozidla), $p_{18} = 1$ (při neztížených povětrnostních podmínkách), $p_{19} = 1$ (dobrá viditelnost), $p_{57} = 1$ (řidič byl v dobrém stavu), $p_{58} = 1$ (bez vnějšího ovlivnění), následují kombinace předešlých skutečností. Mezi další skutečnosti, které vyplývají z asociačních pravidel je atribut $p_{11} = 2$ (nepřítomnost alkoholu) a $p_{12_interval} = 501-516$ (nesprávný způsob jízdy).

Consequent	Antecedent	Rule ID	Instances	Support %	Confidence %	Rule Support %	Lift	Deployability
bez_nasledku	p10 = 1	9	10 042	91,524	100,0	91,524	1,0	0,0
bez_nasledku	p18 = 1	8	9 733	88,708	100,0	88,708	1,0	0,0
bez_nasledku	p58 = 1	7	9 145	83,349	100,0	83,349	1,0	0,0
bez_nasledku	p18 = 1 p10 = 1	43	8 884	80,97	100,0	80,97	1,0	0,0
bez_nasledku	p57 = 1	6	8 812	80,314	100,0	80,314	1,0	0,0
bez_nasledku	p57 = 1 p58 = 1	38	8 772	79,949	100,0	79,949	1,0	0,0
bez_nasledku	p58 = 1 p10 = 1	42	8 258	75,264	100,0	75,264	1,0	0,0
bez_nasledku	p19 = 1	5	8 152	74,298	100,0	74,298	1,0	0,0
bez_nasledku	p58 = 1 p18 = 1	41	8 056	73,423	100,0	73,423	1,0	0,0
bez_nasledku	p57 = 1 p10 = 1	40	7 931	72,284	100,0	72,284	1,0	0,0
bez_nasledku	p57 = 1 p58 = 1 p10 = 1	107	7 905	72,047	100,0	72,047	1,0	0,0
bez_nasledku	p16 = 1	3	7 780	70,908	100,0	70,908	1,0	0,0
bez_nasledku	p57 = 1 p18 = 1	39	7 768	70,798	100,0	70,798	1,0	0,0
bez_nasledku	p11 = 2	4	7 765	70,771	100,0	70,771	1,0	0,0
bez_nasledku	p57 = 1 p58 = 1 p18 = 1	106	7 732	70,47	100,0	70,47	1,0	0,0
bez_nasledku	p16 = 1 p18 = 1	27	7 727	70,425	100,0	70,425	1,0	0,0
bez_nasledku	p19 = 1 p10 = 1	37	7 721	70,37	100,0	70,37	1,0	0,0
bez_nasledku	p19 = 1 p18 = 1	36	7 714	70,306	100,0	70,306	1,0	0,0
bez_nasledku	p11 = 2 p10 = 1	33	7 705	70,224	100,0	70,224	1,0	0,0
bez_nasledku	p11 = 2 p58 = 1	31	7 669	69,896	100,0	69,896	1,0	0,0

Obrázek 7 – Asociační pravidla s největší podporou (nehody bez osobních následků)

Zdroj: [vlastní]

5.4.1.1 Vedlejší analýzy

V rámci každé analýzy jsou zkoumány příčiny nehody na bližších skutečnostech, v tomto případě se jedná o dopravní nehody uskutečněné v pracovní dny, o víkendu, s přítomností alkoholu u řidiče, na dálnici a po srážce se zvěří. Každý z problémů je v první řadě vybrán pomocí uzlu Select a poté na vybrané množině jsou aplikována asociační pravidla.

Pracovní dny

Snahou prvních dvou analýz je zjištění, zda existuje nějaký rozdíl mezi nehodami uskutečněnými v pracovní dny a nehodami o víkendu.

Pro vybraných 7 966 dopravních nehod uskutečněných v pracovní dny bylo vytvořeno 338 asociačních pravidel. Po jejich analyzování byly odhaleny stejné příčiny jako v hlavním problému. Pravidla s nejvyšší podporou jsou atribut p10 = 1 (nehoda způsobena řidičem

motorového vozidla) s podporou 91,75 %, $p_{18} = 1$ (při neztížených povětrnostních podmínkách) s podporou 88,99 % a $p_{58} = 1$ (bez vnějšího ovlivnění) s podporou 83,54 %.

Víkend

O víkendu za sledované období se uskutečnilo 3 006 dopravní nehod, pro které bylo vytvořeno 329 asociačních pravidel. Ty jsou totožná s asociačními pravidly v pracovní den, rozdíl je pouze v jejich podpoře. Pravidla s nejvyšší podporou jsou atribut $p_{10} = 1$ (nehoda způsobena řidičem motorového vozidla) s podporou 90,91 %, $p_{18} = 1$ (při neztížených povětrnostních podmínkách) s podporou 87,95 % a $p_{58} = 1$ (bez vnějšího ovlivnění) s podporou 82,83 %.

Z výše popsanych skutečností lze vidět, že není žádný rozdíl mezi nehoda uskutečněnými v pracovní dny a o víkendu.

Přítomnost alkoholu

Cílem analýzy je zjištění, které faktory vedou k dopravní nehodě bez osobních následků v přítomnosti alkoholu u řidiče.

Přítomnost alkoholu byla zjištěna v 266 případech, k těm je vytvořeno 64 asociačních pravidel. Nejvyšší podporu dosahuje atribut $p_{58} = 1$ (bez vnějšího ovlivnění) s podporou 98,49 %, $p_{10} = 1$ (nehoda způsobena řidičem motorového vozidla) s podporou 93,23 % a kombinace těchto atribut s podporou 92,48 %. Z dalších pravidel je za příčinu nehodu zjištěn nesprávný způsob jízdy a vysoká rychlost způsobené řidičem motorového vozidla.

Dálnice

Dálnice jsou vybrány z důsledku, že se jedná převážně o přehlednou část komunikace a tedy dopravní nehody na ní vzniklé by se měly odlišovat od hlavního problému.

Jedná se o 98 dopravních metod s 88 vytvořenými asociačními pravidly. Dosažené skutečnosti jsou obdobné s hlavním řešeným problémem, a to atribut $p_{58} = 1$ (bez vnějšího ovlivnění) s podporou 83,67 %, p_{57} (řidič v dobrém stavu) s podporou 78,57 %. Rozdíl oproti hlavnímu problému je ten, že k nehodě často dochází v důsledku nárazu do patníku či odrazníku ($p_6 = 3$ s podporou 63,26 %) bez užití alkoholu s podporou 59,18 %.

Srážka se zvěří

Jelikož v 92 % dochází k zavinění dopravní nehody řidičem motorového vozidla, jsou zkoumány příčiny nehody druhou nejčastější možností, která způsobila dopravní nehodu.

Je jí srážka se zvířeti, ke které došlo v 707 případech, ke kterým bylo vytvořeno 317 asociačních pravidel. K takovéto nehodě dochází u řidiče v dobrém stavu $p57 = 1$ (99,15 %), který nebyl ovlivněn vnějším prostředím $58 = 1$ (98,02 %) a nezpůsobil dopravní nehodu $p12_interval = 100$ (100 %).

Vytvořený model a všechny vytvořená asociační pravidla jsou uloženy jako DN_bez_nasledku.srt na příloženém CD.

5.4.2 Nehody s osobními následky – lehká zranění

Další z řešených problémů se zabývá odhalením faktorů, které vedou k dopravní nehodě, při níž došlo k lehkým osobním následkům na zdraví účastníků silničního provozu. Pomocí uzlu Derive je vytvořen nový atribut lehce_zraneni, který jako pravdu hodnotí záznamy s lehce zraněnými osobami při dopravní nehodě. Následně jsou pomocí uzlu Select vybrány pouze záznamy splňující pravdu. Předpoklady zůstanou stejné s předchozím problémem, a to atributy $p6$, $p10$, $p11$, $p12_interval$, $p16$, $p18$, $p19$, $p57$ a $p58$. K těmto atributům jsou přiřazeny dále atribut $p7$ (druh srážky jedoucích vozidel) a $p8$ (druh pevné překážky), které by mohly odhalit důvod osobních následků. Ostatní nevybrané atributy jsou odfiltrovány. Připojen je uzel Type, ve kterém jsou nastaveny vstupní atributy a výstupní atributy. Po nastavení uzlu Apriori jsou vytvořena asociační pravidla pro obě sledovaná období.

Ve sledovaném období došlo v okrese ke 2 406 dopravním nehodám. Vytvořeno bylo 264 asociačních pravidel, podle kterých je dopravní nehoda s následkem lehkého zranění nejčastěji způsobena podobně jako v předešlé analýze řidičem motorového vozidla $p10 = 1$, v dobrém stavu $p57 = 1$, bez užití alkoholu $p11 = 2$, bez vnějšího ovlivnění $p58 = 1$ a při neztížených povětrnostních podmínkách $p18 = 1$. Rozdílem je nově přidaný $p8 = 0$, kdy se nejedná o srážku s pevnou překážkou. Následují kombinace těchto faktorů. Z hlavních příčin u atributu $p12_interval$ se žádný neprojevil. Pravidla s nejvyšší podporou jsou zobrazena na obrázku 8.

Consequent	Antecedent	Rule ID	Instances	Support %	Confidence %	Rule Support %	Lift	Deployability
lehce_zraneni	p58 = 1	10	2 239	93,059	100,0	93,059	1,0	0,0
lehce_zraneni	p18 = 1	9	2 131	88,57	100,0	88,57	1,0	0,0
lehce_zraneni	p8 = 0	8	2 068	85,952	100,0	85,952	1,0	0,0
lehce_zraneni	p57 = 1	7	2 025	84,165	100,0	84,165	1,0	0,0
lehce_zraneni	p57 = 1 p58 = 1	44	2 003	83,25	100,0	83,25	1,0	0,0
lehce_zraneni	p18 = 1 p58 = 1	47	1 978	82,211	100,0	82,211	1,0	0,0
lehce_zraneni	p11 = 2	6	1 920	79,8	100,0	79,8	1,0	0,0
lehce_zraneni	p8 = 0 p58 = 1	46	1 913	79,51	100,0	79,51	1,0	0,0
lehce_zraneni	p11 = 2 p58 = 1	41	1 892	78,637	100,0	78,637	1,0	0,0
lehce_zraneni	p10 = 1	5	1 890	78,554	100,0	78,554	1,0	0,0
lehce_zraneni	p11 = 2 p57 = 1	38	1 882	78,221	100,0	78,221	1,0	0,0
lehce_zraneni	p11 = 2 p57 = 1 p58 = 1	115	1 863	77,431	100,0	77,431	1,0	0,0
lehce_zraneni	p8 = 0 p18 = 1	45	1 844	76,642	100,0	76,642	1,0	0,0
lehce_zraneni	p57 = 1 p18 = 1	43	1 785	74,19	100,0	74,19	1,0	0,0
lehce_zraneni	p57 = 1 p8 = 0	42	1 774	73,732	100,0	73,732	1,0	0,0
lehce_zraneni	p10 = 1 p58 = 1	37	1 766	73,4	100,0	73,4	1,0	0,0
lehce_zraneni	p57 = 1 p18 = 1 p58 = 1	121	1 765	73,358	100,0	73,358	1,0	0,0
lehce_zraneni	p16 = 1	4	1 757	73,026	100,0	73,026	1,0	0,0
lehce_zraneni	p57 = 1 p8 = 0 p58 = 1	120	1 753	72,86	100,0	72,86	1,0	0,0
lehce_zraneni	p19 = 1	3	1 745	72,527	100,0	72,527	1,0	0,0

Obrázek 8 – Asociační pravidla s největší podporou (nehody s lehkým zraněním)

Zdroj: [vlastní]

5.4.2.1 Vedlejší analýzy

Jako vedlejší analýzy jsou zkoumány dopravní nehody uskutečněné o víkendu a během dne, kdy den je rozdělen po 6 hodinách.

Víkend

Cílem analýzy je zjistit, zda se dopravní nehody uskutečněné o víkendu liší od celé zkoumané množiny.

O víkendu se stalo 679 dopravních nehod, uzel Apriori vytvořil 215 asociačních pravidel, které korespondují s hlavním řešeným problémem a faktory vedoucí k dopravní nehodě jsou shodné. Nejvyšší podporu dosahuje atribut p58 = 1 (bez vnějšího ovlivnění) s podporou 94,1 % a p18 = 1 (při neztížených povětrnostních podmínkách) s podporou 89,93 %.

00:00-05:59

Snahou zbylých analýz je zjistit, zda je během dne rozdíl v příčinách dopravních nehod.

Jde o 178 dopravní nehod, ke kterým bylo vytvořeno 83 asociačních pravidel. Nejvyšší podporu dosahují atribut $p58 = 1$ (bez vnějšího ovlivnění) s podporou 94,94 %, $p10 = 1$ (nehoda způsobena řidičem motorového vozidla) s podporou 83,7 % a $p18 = 1$ (při neztížených povětrnostních podmínkách) s podporou 80,89 %. Dále se objevuje atribut $p12_interval = 201-209$ (rychlá jízda) s podporou 41,01 %.

06:00-11:59

V tomto případě jde o 731 nehod s 258 asociačními pravidly. Vytvořená pravidla jsou podobná těm z předchozí analýzy, avšak s rozdílnými hodnotami podpory. Nejčastěji se objevu atribut $p57$ (řidič v dobrém stavu) s podporou 93,98 % a $p58$ (bez vnějšího ovlivnění) s podporou 90,97 %. V pravidlech se objevil atribut $p12_interval = 501-516$ (nesprávný způsob řízení vozidla) s podporou 40,9 %.

12:00-17:59

V tomto časovém úseku došlo k 995 dopravním nehodám s 243 vytvořenými asociačními pravidly. Nejvyšší podporu dosáhly atributy $p58$ (bez vnějšího ovlivnění) s podporou 92,86 %, $p18 = 1$ (při neztížených povětrnostních podmínkách) s podporou 91,75 % a $p19 = 1$ (ve dne, viditelnost nezhoršená) s podporou 89,34 %. Z hlavních příčin atribut $p12_interval$ se objevil nesprávný způsob řízení s podporou 39,69 %.

18:00-23:59

K 552 dopravním nehodám bylo vytvořeno 96 asociačních pravidel. Nejvyšší podporu dosáhly atributy $p58$ (bez vnějšího ovlivnění) s podporou 91,43 %, $p18 = 1$ (při neztížených povětrnostních podmínkách) s podporou 86,65 % a z hlavních příčin atribut $p12_interval = 501-516$ (nesprávný způsob řízení vozidla) s podporou 37,05 %.

Z předešlých analýz je vidět značná spojitost s hlavním problémem, především u atribut $p19$, $p18$, $p57$ a $p58$. Z analýz však vyplývá, že v ranních hodinách (00:00-06:00) je častou příčinou nehody rychlá jízda, zatímco ve zbytku dne je o nesprávný způsob řízení.

Vytvořený model a příslušná asociační pravidla jsou uloženy jako DN_lehce_zraneni.srt na příloženém CD.

5.4.3 Nehody s osobními následky – těžká zranění

Snahou tohoto problému je odhalit faktory, které vedou k dopravní nehodě, při níž došlo k těžkým osobním následkům na zdraví účastníků silničního provozu. Pomocí uzlu Derive je vytvořen nový atribut `tezce_zraneni`, který jako pravdu hodnotí záznamy s těžce zraněnými osobami při dopravní nehodě. Poté jsou pomocí uzlu Select vybrány pouze záznamy splňující pravdu. Předpoklady pro srovnání s předchozí analýzou jsou totožné, a to atributy `p6`, `p7`, `p8`, `p10`, `p11`, `p12_interval`, `p16`, `p18`, `p19`, `p57` a `p58`. Ostatní nevybrané atributy jsou odfiltrovány. Připojen je uzel Type, ve kterém jsou nastaveny vstupní atributy a výstupní atributy. Nastaven je také uzel Apriori, jeho závěr a předpoklady.

Analyzováno bylo 263 dopravní nehod s následkem těžkého zranění, ke kterým uzel Apriori našel 249 asociačních pravidel. Nejčastěji je dopravní nehoda s následkem těžkého zranění způsobena podobně jako v předešlé analýze řidičem motorového vozidla `p10 = 1`, kdy se nejedná o srážku s pevnou překázkou `p8 = 0`, řidičem v dobrém stavu `p57 = 1`, bez užití alkoholu `p11 = 2`, bez vnějšího ovlivnění řidiče `p58 = 1` a při neztížených povětrnostních podmínkách `p18 = 1`. Stejně jako v předešlém případě se neprojevil vliv žádné hlavní příčiny. Pravidla s nejvyšší podporou jsou znázorněna na obrázku 9.

Consequent	Antecedent	Rule ID	Instances	Support %	Confidence %	Rule Support %	Lift	Deployability
tezce_zraneni	p58 = 1	10	256	97,338	100,0	97,338	1,0	0,0
tezce_zraneni	p57 = 1	9	235	89,354	100,0	89,354	1,0	0,0
tezce_zraneni	p57 = 1 p58 = 1	49	235	89,354	100,0	89,354	1,0	0,0
tezce_zraneni	p18 = 1	8	229	87,072	100,0	87,072	1,0	0,0
tezce_zraneni	p18 = 1 p58 = 1	48	224	85,171	100,0	85,171	1,0	0,0
tezce_zraneni	p11 = 2	7	220	83,65	100,0	83,65	1,0	0,0
tezce_zraneni	p11 = 2 p58 = 1	46	219	83,27	100,0	83,27	1,0	0,0
tezce_zraneni	p10 = 1	6	217	82,51	100,0	82,51	1,0	0,0
tezce_zraneni	p11 = 2 p57 = 1	45	215	81,749	100,0	81,749	1,0	0,0
tezce_zraneni	p11 = 2 p57 = 1 p58 = 1	118	215	81,749	100,0	81,749	1,0	0,0
tezce_zraneni	p10 = 1 p58 = 1	43	212	80,608	100,0	80,608	1,0	0,0
tezce_zraneni	p8 = 0	5	211	80,228	100,0	80,228	1,0	0,0
tezce_zraneni	p18 = 1 p57 = 1	47	206	78,327	100,0	78,327	1,0	0,0
tezce_zraneni	p18 = 1 p57 = 1 p58 = 1	119	206	78,327	100,0	78,327	1,0	0,0
tezce_zraneni	p8 = 0 p58 = 1	39	205	77,947	100,0	77,947	1,0	0,0
tezce_zraneni	p16 = 1	4	196	74,525	100,0	74,525	1,0	0,0
tezce_zraneni	p10 = 1 p57 = 1	42	196	74,525	100,0	74,525	1,0	0,0
tezce_zraneni	p10 = 1 p57 = 1 p58 = 1	115	196	74,525	100,0	74,525	1,0	0,0
tezce_zraneni	p16 = 1 p18 = 1	32	194	73,764	100,0	73,764	1,0	0,0

Obrázek 9 – Asociační pravidla s největší podporou (nehody s těžkým zraněním)

Zdroj: [vlastní]

5.4.3.1 Vedlejší analýzy

Vedlejšími analýzami jsou určeny dopravní nehody uskutečněné v pátek, na silnicích 1., 2. a 3. třídy, při střetu s pevnou překážkou, za mokra a při nepřiměřené rychlosti jízdy.

Pátek

Konec týdne bývá pro většinu lidí hektický, proto se předpokládá, že důvodem dopravní nehody v pátek by měl být odlišný od zbylých dnů.

Jde o 30 případů se 174 asociačními pravidly, kdy výsledky korespondují s hlavní analýzou. Nejvyšší podporu dosáhly atributy p58 (bez vnějšího ovlivnění) s podporou 96,66 %, p18 = 1 (při neztížených povětrnostních podmínkách) s podporou 83,33 % a p57 = 1 (řidič v dobrém stavu) s podporou 83,33 %. Z hlavních příčin se projevil atribut p12_interval = 501-516 (nesprávný způsob řízení vozidla) s podporou 43,33 % i p12_interval = 201-209

(rychlá jízda) s podporou 30 %. Rozdílem oproti hlavnímu problému je přítomnost atributu p12_interval.

Silnice 1., 2. a 3. třídy

Vybrány jsou tyto třídy silnic především kvůli horšímu stavu a menší přehlednosti než jsou dálnice.

Analyzováno je 212 dopravních nehod, ke kterým je vytvořeno 229 asociačních pravidel. Nejvyšší podporu dosahují atribut p58 = 1 (bez vnějšího ovlivnění) s podporou 98,11 %, p10 = 1 (nehoda způsobena řidičem motorového vozidla) s podporou 89,15 % a p57 = 1 (řidič v dobrém stavu) s podporou 88,69 %. Vytvořená analýza se nijak neliší od hlavní analýzy.

Střet s pevnou překážkou

Předpokladem této analýzy je, že k těžkému zranění dochází při střetu s pevnou překážkou.

Ke střetu s pevnou překážkou došlo v 52 případech a vytvořeno bylo 200 asociačních pravidel. Nejvyšší podporu dosahují atribut p10 = 1 (nehoda způsobena řidičem motorového vozidla) s podporou 98,07 % a p58 = 1 (bez vnějšího ovlivnění) s podporou 98,07 % a jejich kombinace. Významná je také nepřítomnost alkoholu p11 = 2 s podporou 76,92 % a vysoká rychlost řidičů p12_interval = 201-209 s podporou 61,53 %.

Mokro

Stejně jako v předchozí analýze se předpokládá, že k těžkému zranění může dojít za mokra.

Za mokra došlo k 46 dopravním nehodám s těžkým zraněním, uzel Apriori našel 37 asociačních pravidel. K dopravním nehodám za mokra dochází bez vnějšího ovlivnění řidiče s podporou 95,65 % a řidičem v dobrém stavu s podporou 86,95 % s nepřítomnosti alkoholu s podporou 82,6 %. Důvodem nehody je také nesprávný způsob jízdy s podporou 34,78 %.

Nepřiměřená rychlost

Posledním předpokladem důvodu vzniku dopravní nehody s následkem těžkého zranění je nepřiměřená rychlost.

Těžké zranění v důsledku nepřiměřené rychlosti se stalo v 84 případech, ke kterým bylo vytvořeno 70 asociačních pravidel. Nejčastější atributy jsou shodné s předešlými, a to p58 = 1 s podporou 100 %, p57 = 1 s podporou 92,85 %. Zajímavým je přítomnost atribut p6 = 3 s podporou 38,09 %.

Vytvořený model a asociační pravidla jsou uloženy jako DN_tezce_zraneni.srt na příloženém CD.

5.4.4 Nehody s osobními následky – usmrcení do 24 hodin

Snahou tohoto problému je odhalit faktory, které vedou k dopravní nehodě, při níž došlo k úmrtí některého z účastníků silničního provozu. Pomocí uzlu Derive je vytvořen nový atribut úmrtí, který jako pravdu hodnotí záznamy s usmrcením při dopravní nehodě. Poté jsou pomocí uzlu Select vybrány pouze záznamy splňující pravdu. Předpoklady pro srovnání s předchozími analýzami jsou totožné, a to atributy p6, p7, p8, p10, p11, p12_interval, p16, p18, p19, p57 a p58. Ostatní nevybrané atributy jsou odfiltrovány. Připojen je uzel Type, ve kterém jsou nastaveny vstupní atributy a výstupní atributy. Nastaven je také uzel Apriori, jeho závěr a předpoklady.

Za sledované období se uskutečnilo 90 dopravních nehod s následkem usmrcení do 24 hodin od uskutečnění nehody. Vytvořeno bylo 142 asociačních pravidel, které jsou podobné předcházejícím analýzám. Nejčastěji je dopravní nehoda s následkem usmrcení způsobena řidičem v dobrém stavu $p57 = 1$, bez vnějšího ovlivnění $p58 = 1$, řidičem motorového vozidla $p10 = 1$, při neztížených povětrnostních podmínkách $p18 = 1$, vozovce v dobrém stavu $p16 = 1$ a nejedná se o střet s pevnou překážkou $p8 = 0$. Vytvořená pravidla s nejvyšší podporou jsou zobrazena na obrázku 10.

Consequent	Antecedent	Rule ID	Instances	Support %	Confidence %	Rule Support %	Lift	Deployability
usmrceni	p58 = 1	10	87	96,667	100,0	96,667	1,0	0,0
usmrceni	p18 = 1	9	80	88,889	100,0	88,889	1,0	0,0
usmrceni	p18 = 1 p58 = 1	43	77	85,556	100,0	85,556	1,0	0,0
usmrceni	p10 = 1	8	73	81,111	100,0	81,111	1,0	0,0
usmrceni	p16 = 1	7	72	80,0	100,0	80,0	1,0	0,0
usmrceni	p10 = 1 p58 = 1	42	71	78,889	100,0	78,889	1,0	0,0
usmrceni	p57 = 1	6	69	76,667	100,0	76,667	1,0	0,0
usmrceni	p16 = 1 p18 = 1	39	69	76,667	100,0	76,667	1,0	0,0
usmrceni	p16 = 1 p58 = 1	40	69	76,667	100,0	76,667	1,0	0,0
usmrceni	p8 = 0	5	68	75,556	100,0	75,556	1,0	0,0
usmrceni	p57 = 1 p58 = 1	37	68	75,556	100,0	75,556	1,0	0,0
usmrceni	p8 = 0 p58 = 1	33	66	73,333	100,0	73,333	1,0	0,0
usmrceni	p16 = 1 p18 = 1 p58 = 1	90	66	73,333	100,0	73,333	1,0	0,0
usmrceni	p11 = 2	4	65	72,222	100,0	72,222	1,0	0,0
usmrceni	p10 = 1 p18 = 1	41	65	72,222	100,0	72,222	1,0	0,0
usmrceni	p11 = 2 p58 = 1	28	64	71,111	100,0	71,111	1,0	0,0
usmrceni	p10 = 1 p18 = 1 p58 = 1	91	63	70,0	100,0	70,0	1,0	0,0
usmrceni	p11 = 2 p57 = 1	24	62	68,889	100,0	68,889	1,0	0,0
usmrceni	p11 = 2 p57 = 1 p58 = 1	65	61	67,778	100,0	67,778	1,0	0,0

Obrázek 10 – Asociační pravidla s největší podporou (nehody s usmrcením do 24 hodin)

Zdroj: [vlastní]

5.4.4.1 Vedlejší analýzy

Jako vedlejší analýzy jsou určeny dopravní nehody uskutečněné v pracovní týden, v ranních hodinách, při nepřiměřené rychlosti a bez užití alkoholu.

Pracovní týden

Předpokladem analýzy je především uspěchanost dojíždění do práce a projevit by se měla vysoká rychlost.

V pracovním týdnu se uskutečnilo 58 dopravních nehod s následkem úmrtí, vytvořeno bylo 142 asociačních pravidel. Nejvyšší podporu dosahuje atribut p58 = 1 (bez vnějšího ovlivnění) s podporou 96,55 %, p18 = 1 (neztížené povětrnostní podmínky) s podporou 91,37 % a p16 = 1 (suchý, neznečištěný povrch) s podporou 81,03 %. Dosažené výsledky zcela

korespondují s hlavní analýzou. Předpokládaná rychlost se projevila pouze s podporou 31,03 %.

Ranní hodiny

Analýza předpokládá, že v ranních hodinách (00:00-06:00) by mohlo docházet k odlišným příčinám oproti hlavní analýze. A to z důvodu horší viditelnosti.

Jde o 22 dopravních nehod s 205 asociačními pravidly. Nejvyšší podporu dosáhly atributy p10 = 1 (řidičem motorového vozidla) s podporou 100 %, p58 = 1 (bez vnějšího ovlivnění) s podporou 95,45% a p16 = 1 (suchý, neznečištěný povrch) s podporou 86,36 %. S podporou 68,18 % se dopravní nehody uskutečnily v noci a s podporou 63,63 % je příčinou rychlá jízda.

Nepřiměřená rychlost

Předchozí analýza odhalila, že přes 63 procent dopravních nehod, jejichž následkem je úmrtí je zapříčiněna rychlou jízdou, proto je rychlá jízda analyzována.

V 33 případech došlo k úmrtí v důsledku nepřiměřené rychlosti, uzel Apriori vytvořil 63 asociačních pravidel. Nejvyšší podporu dosáhl atribut p58 = 1 s podporou 100 %, dále p 57 = 1 s podporou 84,84 % a p18 = 1 s podporou 78,78 %. Je zde vidět provázanost s předcházející analýzou, protože zde má vliv také atribut p19 = 6 (v noci) s podporou 48,48 % a p6 = 3 (srážka s pevnou překážkou) s podporou 45,45 %.

Vytvořený model a příslušná asociační pravidla jsou uložena jako DN_usmrceni.srt na příloženém CD.

5.4.5 Velikost hmotné škody

Tento problém je zaměřen na nalezení asociačních pravidel v závislosti na velikost hmotné škody. Atribut vypovídající o velikosti hmotné škody je atribut p14. Hmotná škoda byla rozdělena pomocí uzlu Binning na 3 části ve skoro stejném poměru dopravních nehod na velikosti hmotné škody. Hranice byly vytvořeny od hodnot 0 až 14,9 tisíc korun, dále od 15 tisíc do 44,9 tisíc korun a naposled od hodnoty 45 tisíc a výše. Uzel Derive byl použit u každé skupiny velikosti hmotné škody na rozkategorizování hodnot na True and False, kde po-té v Selectu jsou vybrány pouze hodnoty True. Nově vytvořené atributy jsou pojmenovány Nizka_skoda, Stredni_skoda a Vysoka_skoda. Pomocí uzlu Filter jsou vyfiltrovány atributy nesouvisející s problémem. Žádoucí v tomto případě jsou atributy ovlivňující hmotnou škodu při nehodě, jsou to atributy vypovídající především o druhu nehody, střetu s jiným vozidlem či překážkou. Jako takové jsou vybrány atributy p6 (druh nehody), p7 (druh srážky jedoucích

vozidel), p8 (druh střetu s překážkou), p16 (stav povrchu vozovky), a p19 (viditelnost). Tyto atributy jsou v uzlu Type nastaveny jako vstupní a nově vzniklý atribut Derive jako výstupní. V uzlu Apriori jsou nastaveny předpoklady a závěr.

Za sledované období se uskutečnilo 4192 dopravních nehod s hmotnou škodou do 14,5 tisíce korun. Vytvořeno bylo 66 asociačních pravidel a nejčastěji je dopravní nehoda s vozidlem zaparkovaným p6 = 2, srážka nejedoucích vozidel p7 = 0, nejedná se o nehodu s pevnou překážkou p8 = 0, povrch suchý a neznečištěný p16 = 1 a ve dne viditelnost nezhoršená p19 = 1. Vše je zobrazeno na obrázku 10.

Consequent	Antecedent	Rule ID	Instances	Support %	Confidence %	Rule Support %	Lift	Deployability
Nizka_skoda	p8 = 0	12	3 652	87,118	100,0	87,118	1,0	0,0
Nizka_skoda	p16 = 1	11	3 165	75,501	100,0	75,501	1,0	0,0
Nizka_skoda	p19 = 1	10	3 134	74,761	100,0	74,761	1,0	0,0
Nizka_skoda	p16 = 1 p8 = 0	37	2 797	66,722	100,0	66,722	1,0	0,0
Nizka_skoda	p19 = 1 p8 = 0	36	2 790	66,555	100,0	66,555	1,0	0,0
Nizka_skoda	p7 = 0	9	2 565	61,188	100,0	61,188	1,0	0,0
Nizka_skoda	p19 = 1 p16 = 1	35	2 564	61,164	100,0	61,164	1,0	0,0
Nizka_skoda	p19 = 1 p16 = 1 p8 = 0	57	2 303	54,938	100,0	54,938	1,0	0,0
Nizka_skoda	p7 = 0 p8 = 0	34	2 025	48,306	100,0	48,306	1,0	0,0
Nizka_skoda	p7 = 0 p16 = 1	33	1 927	45,969	100,0	45,969	1,0	0,0
Nizka_skoda	p7 = 0 p19 = 1	32	1 820	43,416	100,0	43,416	1,0	0,0
Nizka_skoda	p6 = 1	8	1 627	38,812	100,0	38,812	1,0	0,0
Nizka_skoda	p6 = 1 p8 = 0	31	1 627	38,812	100,0	38,812	1,0	0,0
Nizka_skoda	p7 = 0 p16 = 1 p8 = 0	56	1 559	37,19	100,0	37,19	1,0	0,0
Nizka_skoda	p7 = 0 p19 = 1 p16 = 1	54	1 478	35,258	100,0	35,258	1,0	0,0
Nizka_skoda	p7 = 0 p19 = 1 p8 = 0	55	1 476	35,21	100,0	35,21	1,0	0,0
Nizka_skoda	p6 = 1 p19 = 1	29	1 314	31,345	100,0	31,345	1,0	0,0

Obrázek 11 – Asociační pravidla s největší podporou (hmotné škody menší než 15 tisíc)

Zdroj: [vlastní]

Ve škodě od 15 tisíc do 44,9 tisíc korun je za sledované období uskutečněno 4639 dopravních nehod. Vzniklo 87 asociačních pravidel a nejčastěji se jedná o dopravní nehodu s jedoucím nekolejovým vozidlem p6 = 1, druh srážky zezadu p7 = 4, nejedná se o náraz do pevné překážky p8 = 0, povrch suchý a neznečištěný p16 = 1 a srážka byla ve dne, kdy byla viditelnost nezhoršená p19 = 1. Pravidla jsou zobrazena na obrázku 11.

Consequent	Antecedent	Rule ID	Instances	Support %	Confidence %	Rule Support %	Lift	Deployability
Stredni_skoda	p8 = 0	11	3 993	86,075	100,0	86,075	1,0	0,0
Stredni_skoda	p19 = 1	10	3 454	74,456	100,0	74,456	1,0	0,0
Stredni_skoda	p16 = 1	9	3 298	71,093	100,0	71,093	1,0	0,0
Stredni_skoda	p19 = 1 p8 = 0	40	3 077	66,329	100,0	66,329	1,0	0,0
Stredni_skoda	p16 = 1 p8 = 0	39	2 894	62,384	100,0	62,384	1,0	0,0
Stredni_skoda	p16 = 1 p19 = 1	38	2 684	57,857	100,0	57,857	1,0	0,0
Stredni_skoda	p6 = 1	8	2 659	57,318	100,0	57,318	1,0	0,0
Stredni_skoda	p6 = 1 p8 = 0	37	2 659	57,318	100,0	57,318	1,0	0,0
Stredni_skoda	p16 = 1 p19 = 1 p8 = 0	71	2 409	51,929	100,0	51,929	1,0	0,0
Stredni_skoda	p6 = 1 p19 = 1	36	2 216	47,769	100,0	47,769	1,0	0,0
Stredni_skoda	p6 = 1 p19 = 1 p8 = 0	70	2 216	47,769	100,0	47,769	1,0	0,0
Stredni_skoda	p7 = 0	7	1 979	42,66	100,0	42,66	1,0	0,0
Stredni_skoda	p6 = 1 p16 = 1	35	1 911	41,194	100,0	41,194	1,0	0,0
Stredni_skoda	p6 = 1 p16 = 1 p8 = 0	69	1 911	41,194	100,0	41,194	1,0	0,0
Stredni_skoda	p6 = 1 p16 = 1 p19 = 1	68	1 736	37,422	100,0	37,422	1,0	0,0
Stredni_skoda	p6 = 1 p16 = 1 p19 = 1 p8 = 0	85	1 736	37,422	100,0	37,422	1,0	0,0
Stredni_skoda	p7 = 0 p16 = 1	32	1 386	29,877	100,0	29,877	1,0	0,0

Obrázek 12 – Asociační pravidla s největší podporou (hmotné škody v rozmezí 15–45 tisíc)

Zdroj: [vlastní]

V poslední řadě jsou nehody od 45 tisíc a výše, kde za sledované období se uskutečnilo 4 787 dopravních nehod a dále vzniklo 69 asociačních pravidel, kde nejčastěji dopravní nehoda vzniká stejně jako v předchozí analýze a to s jedoucím nekolejovým vozidlem p6 = 1, druh srážky zezadu p7 = 4, nejedná se o náraz do pevné překážky p8 = 0, povrch suchý a neznečištěný p16 = 1 a srážka byla ve dne, kdy byla viditelnost nezhoršená p19 = 1. Vše je zobrazeno na obrázku 12 a vytvořený model je uložen jako DN_hmotna_skoda.srt na příloženém CD.

Consequent	Antecedent	Rule ID	Instances	Support %	Confidence %	Rule Support %	Lift	Deployability
Vysoka_skoda	p8 = 0	10	4 025	84,082	100,0	84,082	1,0	0,0
Vysoka_skoda	p19 = 1	9	3 464	72,363	100,0	72,363	1,0	0,0
Vysoka_skoda	p16 = 1	8	3 257	68,038	100,0	68,038	1,0	0,0
Vysoka_skoda	p6 = 1	7	3 243	67,746	100,0	67,746	1,0	0,0
Vysoka_skoda	p6 = 1 p8 = 0	31	3 243	67,746	100,0	67,746	1,0	0,0
Vysoka_skoda	p19 = 1 p8 = 0	34	3 049	63,693	100,0	63,693	1,0	0,0
Vysoka_skoda	p16 = 1 p8 = 0	33	2 839	59,306	100,0	59,306	1,0	0,0
Vysoka_skoda	p16 = 1 p19 = 1	32	2 634	55,024	100,0	55,024	1,0	0,0
Vysoka_skoda	p6 = 1 p19 = 1	30	2 596	54,23	100,0	54,23	1,0	0,0
Vysoka_skoda	p6 = 1 p19 = 1 p8 = 0	56	2 596	54,23	100,0	54,23	1,0	0,0
Vysoka_skoda	p16 = 1 p19 = 1 p8 = 0	57	2 360	49,3	100,0	49,3	1,0	0,0
Vysoka_skoda	p6 = 1 p16 = 1	29	2 330	48,673	100,0	48,673	1,0	0,0
Vysoka_skoda	p6 = 1 p16 = 1 p8 = 0	55	2 330	48,673	100,0	48,673	1,0	0,0
Vysoka_skoda	p6 = 1 p16 = 1 p19 = 1	54	2 030	42,407	100,0	42,407	1,0	0,0
Vysoka_skoda	p6 = 1 p16 = 1 p19 = 1 p8 = 0	67	2 030	42,407	100,0	42,407	1,0	0,0
Vysoka_skoda	p7 = 0	6	1 538	32,129	100,0	32,129	1,0	0,0
Vysoka_skoda	p7 = 4	5	1 355	28,306	100,0	28,306	1,0	0,0

Obrázek 13 – Asociační pravidla s největší podporou (hmotné škody od 45 tisíc a výše)

Zdroj: [vlastní]

5.5 Vyhodnocení výsledků

Pokud by se obecně daly dohromady společné faktory z řešených problémů vedoucí k dopravní nehodě, jednalo by se o dopravní nehodu zaviněnou řidičem osobního vozidla v dobrém stavu, bez užití alkoholu a při neztížených vnějších podmínkách. To jsou hlavní faktory, které jsou společné u všech problémů.

První z řešených problémů, dopravní nehoda bez osobních následků má kromě zmíněných faktorů jako podstatnou příčinu především nesprávný způsob jízdy. Analyzované dílčí problémy korespondují s hlavní problémem až na dopravní nehody uskutečněné na dálnici, které byly způsobeny především nárazem do patníku či odrazníku. Druhá analýza, dopravní nehoda s lehkým zraněním plně koresponduje s popsány příčinami na začátku kapitoly, dále se také nejedná o srážku s pevnou překážkou. Dílčí analýzy odhalily, že v ranních hodinách dochází často k rychlé jízdě, zatímco ve zbytku dne dochází k nesprávnému

způsobu řízení. V pořadí třetí analýzy, dopravní nehoda s těžkým zraněním je stejná jako nehoda s lehkým zraněním. Dílčí analýzy odhalily skutečnost, že střet s pevnou překážkou je zapříčiněn vysokou rychlostí a naopak. Dále bylo zjištěno, že dopravní nehoda za mokra je způsobena buď vysokou rychlostí nebo nesprávných způsobem řízení. Další analýza, dopravní nehoda s usmrcením do 24 hodin je způsobena stejnými vlivy jako předcházející analýzy. Dílčí analýzy však odhalily skutečnost, že ke smrtelným nehodám dochází především vlivem špatné viditelnosti a rychlé jízdy. Analýza hmotné škody ukázala, že dopravní nehody vznikají nejčastěji s jedoucím nekolejovým vozidlem na suché vozovce a s nárazem zezadu. Dílčí analýzy k této problematice už nebyly dále tvořeny.

ZÁVĚR

s Cílem bakalářské práce byla analýza příčin a následků dopravních metod pomocí asociační analýzy na datech o nehodovosti v okrese Pardubice za období 2006–2011.

Modelovány byly dopravní nehodě bez osobních následků, s lehkým zraněním, s těžkým zraněním, s úmrtím do 24 hodin a velikost hmotné škody. Společnými faktory u všech problémů je uskutečnění dopravní nehody řidičem osobního vozidla v dobrém stavu, bez užití alkoholu a při neztížených vnějších podmínkách. Zjištěno bylo, že dopravní nehody bez osobních následků jsou způsobeny především nesprávným způsobem jízdy. Nehody s lehkými zraněnými nejsou způsobeny srážkou s pevnou překážkou. Nehody s těžkým zraněným jsou způsobeny stejnými faktory jako u lehkých zranění, ale vlivem vysoké rychlosti často dochází ke střetům s pevnou překážkou a nehoda způsobená za mokra je způsobena vysokou rychlostí či nesprávným způsobem řízení. K smrtelným nehodám dochází ve velké části především v důsledku špatné viditelnosti a rychlé jízdy. Hmotná škoda při dopravních nehodách je různorodá. Nejčastěji vznikají dopravní nehody v hodnotě 15 tisíc a výše na suché vozovce při nezhoršené viditelnosti a nárazem zezadu.

Využití získaných skutečností by se dalo použít jako dílčí podklad pro analytickou část regionální strategie bezpečnosti silničního provozu okrese Pardubice.

Bakalářská práce se skládá z pěti hlavních částí. První a druhá část se zabývala teoretickými východisky nehodovosti, kdy byl definován pojem dopravní nehoda. Popsány byly příčiny vedoucí k dopravní nehodě a klasifikace dopravních nehod. Charakterizováno bylo území okrese Pardubice a jeho dopravní síť. Byl zhodnocen vývoj nehodovosti za poslední roky celé České republiky a byla popsána bezpečnost silničního provozu. Třetí část se zabývala teoretickou charakteristikou pojmu data mining a jeho metodik. Čtvrtá část se věnovala obecně asociační analýze a jejím metodám. V poslední části byly řešeny problémy pomocí metodiky CRISP-DM a jejich jednotlivých fází (porozumění problému, porozumění datům, příprava dat, modelování, hodnocení, využití v praxi). Nejprve byly definovány řešené problémy, jejichž následky by se daly shrnout na osobní a finanční. Dále byly popsány data a vytvořena základní analýza o nich samotných. Ve fázi příprava dat, byla data převedena do podoby vhodné pro modelování. Závěrem byly modelovány vybrané problémy a vyhodnoceny dosažené výsledky.

POUŽITÁ LITERATURA

- [1] Alkohol za volantem. Dopravní-právo.cz: pomáhat a chránit! [online]. 2014 [cit. 2014-06-20]. Dostupné z: <<http://www.dopravni-pravo.cz/alkohol-za-volantem/>>.
- [2] BERKA, Petr. Dobývání znalostí z databází. 1. Vyd. Praha, 2003. 366 s. ISBN 80-200-1062-9.
- [3] BURDA, Michal. Získávání znalostí z databází: Asociační pravidla. FIT [online]. 2004 [cit. 2014-06-18]. Dostupné z: <<http://www.fit.vutbr.cz/study/courses/ZZD/public/seminar0304/GUHA-text.pdf>>.
- [4] CRISP-DM model. Bricolage de strategy [online]. 2011 [cit. 2014-06-21]. Dostupné z: <<http://geuder.tumblr.com/post/3028113424/crisp-dm-model-cross-industry-standard-process>>.
- [5] Dobývání znalostí z dat o hypertenzi. EuroMISE [online]. 2002 [cit. 2014-06-21]. Dostupné z: <<http://euromise.vse.cz/kdd/index.php?page=kdd>>.
- [6] HAN, Jiawei, Micheline KAMBER a Jian PEI. Data mining: concepts and techniques. 3rd ed, Academic Press 2011, 744 s. ISBN 978-0-12-381479-1
- [7] Hodnocení bezpečnosti silničního provozu v krajích. Observatoř bezpečnosti silničního provozu [online]. 2011 [cit. 2014-06-24]. Dostupné z: <<http://www.czrso.cz/clanky/hodnoceni-bezpecnosti-silnicniho-provozu-v-krajich/>>.
- [8] Charakteristika okresu Pardubice. Český statistický úřad [online]. 2013 [cit. 2014-06-24]. Dostupné z: <http://www.czso.cz/x/redakce.nsf/i/charakteristika_okresu_pardubice>.
- [9] CHMELÍK, Jan. Dopravní nehody. Plzeň: Vydavatelství a nakladatelství Aleš Čeněk, 2009, 540 s. ISBN 978-80-7380-211-0.
- [10] IBM SPSS Modeler Help. Dostupné z: <http://pic.dhe.ibm.com/infocenter/spssmodl/v15r0m0/index.jsp?topic=%2Fcom.ibm.sps.s.modeler.help%2Fquestnode_general.htm>.
- [11] Kategorie pozemních komunikací dle ČSN. Observatoř bezpečnosti silničního provozu [online]. 2007 [cit. 2014-06-24]. Dostupné z: <<http://www.czrso.cz/clanky/kategorie-pozemnich-komunikaci-dle-csn/>>.
- [12] KOPECKÝ, Zdeněk. Občan a dopravní nehoda. Vyd. 1. Praha: Prospektrum, 1998, 198 s. Právo do kapsy. ISBN 80-717-5068-9.

- [13] KUČEROVÁ, Helena. Dopravní přestupky v praxi: podle stavu k 1.7.2006. 2. přepracované a doplněné vydání. Praha: Linde, 2006, 431 s. ISBN 80-720-1613-X.
- [14] LIŠKA, Miroslav. Dobývání znalostí z databází. OSU [online]. 2008 [cit. 2014-06-17]. Dostupné z: <http://www1.osu.cz/~klimesc/public/files/DOZNA/Dobyvani_znalosti_z_databazi.pdf>.
- [15] Národní strategie bezpečnosti silničního provozu 2011-2020. BESIP [online]. 2012 [cit. 2014-06-24]. Dostupné z: <<http://www.ibesip.cz/cz/besip/strategie-dokumenty/narodni-strategie-bezpecnosti-silnicniho-provozu/nsbsp-2011-2020>>.
- [16] Nehody v silniční dopravě v krajích a okresech. Český statistický úřad [online]. 2014 [cit. 2014-06-24]. Dostupné z: <http://vdb.czso.cz/vdbvo/tabparam.jsp?vo=null&cislotab=DOP0080UU+_KR&voa=tabulka&go_zobraz=1&aktualizuj=Aktualizovat&cas_1_100=2012>.
- [17] OECD, ITF. *Road Safety Annual Report 2011* [online]. 2012 [cit. 2014-06-12]. Dostupné z: <<http://www.internationaltransportforum.org/irtadpublic/pdf/11IrtadReport.pdf>>.
- [18] Pardubice získaly titul Hlavní město cyklistů 2014. NaKole.cz [online]. 2014 [cit. 2014-06-24]. Dostupné z: <<http://www.nakole.cz/clanky/1138-pardubice-ziskaly-titul-hlavni-mesto-cyklistu-2014.html>>.
- [19] PETR, Pavel. Data Mining: Díl I. Pardubice: Univerzita Pardubice, 2006. ISBN 80-7194-886-1.
- [20] PORADA, Viktor. Silniční dopravní nehoda v teorii a praxi. Praha: Linde, 2000, 378 s. Vysokoškolské právnické učebnice. ISBN 80-720-1212-6.
- [21] Pozor na léky za volantem. Česká lékárnická komora [online]. 2012 [cit. 2014-06-20]. Dostupné z: <<http://www.lekarnici.cz/Pro-verejnost/Informace-pro-verejnost/Pozor-na-leky-za-volantem.aspx>>.
- [22] Pravidla silničního provozu. BESIP [online]. 2012 [cit. 2014-06-24]. Dostupné z: <<http://www.ibesip.cz/cz/legislativa/pravidla-silnicniho-provozu>>.
- [23] Přehledy z informačního systému o silniční a dálniční síti ČR. Ředitelství silnic a dálnic ČR [online]. 2014 [cit. 2014-06-24]. Dostupné z: <http://www.rsd.cz/sdb_intranet/sdb/download/prehledy_2014_1_pa.pdf>.
- [24] Ředitelství silnic a dálnic. Dostupné z: <<http://www.rsd.cz/>>.

- [25] Silniční a dopravní síť v ČR. [online]. [cit. 2014-07-05]. Dostupné z: <http://www.rsd.cz/sdb_intranet/sdb/img/mapy/cr_a3_e.png>.
- [26] Univerzita v číslech. Univerzita Pardubice [online]. 2014 [cit. 2014-06-24]. Dostupné z: <<http://www.upce.cz/univerzita/univ-cisla.html>>.
- [27] VOMÁČKA, Petr. Deset nejčastějších příčin nehod. UAMK [online]. 2013 [cit. 2014-06-20]. Dostupné z: <<http://www.uamk.cz/item/1974-desetnej%C4%8Dast%C4%9Bj%C5%A1%C3%ADchp%C5%99%C3%AD%C4%8Din-nehod>>.
- [28] RICHTÁŘ, Michal, Vladislav KŘIVDA a Ivana OLIVKOVÁ. Základy městské a silniční dopravy. Katedra dopravního stavitelství, Fakulta stavební, VŠB-TU Ostrava [online]. [cit. 2014-07-05]. Dostupné z: <<http://kds.vsb.cz/mhd/zaklady-uvod.htm>>.
- [29] Zákon č.13/1997 Sb., o pozemních komunikacích, § 2 „Pozemní komunikace a jejich rozdělení“.

SEZNAM PŘÍLOH

Příloha A: Formulář evidence nehod v silničním provozu

Příloha B: Datový slovník

Příloha C: Charakteristiky kategoriálních atributů

Příloha D: CD s vytvořenými modely a analyzovanými daty

34 POČET ZÚČASTNĚNÝCH VOZIDEL

uvádí se skutečný počet vozidel

35 MÍSTO DOPRAVNÍ NEHODY

- 00 mimo křižovatky
- 10 na křižovatce, jedná-li se o křižování silnic 3. tř., místních, úbočkových komunikací
- 11-16 uvnitř zóny 1-6 přední křižovatky
- 18 na křižovatce, uvnitř hranic křižovatky definovaných pro systémy evidence nehod (zóna B)
- 22-26 na vozovce nebo výjezdové části silnice při manévrování křižování
- 28 mimo zónu 11-18 a 22-26

36 DRUH POZEMNÍ KOMUNIKACE

- 0 dálnice
- 1 silnice 1. třídy
- 2 silnice 2. třídy
- 3 silnice 3. třídy
- 4 zast. (v křižovatce sledování) ve vybraných místech
- 5 komunikace sledování (ve vybraných místech)
- 6 komunikace místní
- 7 komunikace úbočkové - polní a lesní cesty apod.
- 8 komunikace úbočkové - lesní (parkovní apod.)

37 ČÍSLO POZEMNÍ KOMUNIKACE

vyplňuje se čísla

- dálnice - čísla 01 až 99
- silnice 1. tř. - čísla 01 až 99
- silnice 2. tř. - čísla 101 až 999
- silnice 3. tř. - čtyř - šestmístná

38 KILOMETR NEHODY

na dálnici, silnici 1. až 3. třídy (na 2 desetinná místa), měří se před číselným označením ulice

39 DRUH KŘIŽOVACÍ KOMUNIKACE

- 1 silnice 1. třídy
- 2 silnice 2. třídy
- 3 silnice 3. třídy
- 6 místní komunikace
- 7 úbočkové komunikace
- 8 silnice mimoúrovňové křižovatky

40,41 ČÍSLO ÚZLU

uvádí se čtyřmístné číslo sledované křižovatky

44 DRUH VOZIDLA

- 00 motopád
- 01 malý motocykl (do 50 cm³)
- 02 motocykl (větší objemová, skútrů apod.)
- 03 osobní automobil bez přívěsu
- 04 osobní automobil s přívěsem
- 05 nákladní automobil (včetně motokáry, autokaraván, cisterny atd.)
- 06 nákladní automobil s přívěsem
- 07 nákladní automobil s návěsem
- 08 autobus
- 09 tahač (s přívěsem)
- 10 traktor
- 11 traktor
- 12 jiné motorové vozidlo (samokáta, stavební atd.)
- 13 zbytek kolo
- 14 pedál, žula na kolo
- 15 jiné nemotorové vozidlo
- 16 chůze
- 17 neobjasněno, kód upl
- 18 jiný druh vozidla

45a VÝROBNÍ ZNAČKA MOTOROVÉHO VOZIDLA

- | | |
|---------------|-----------------------|
| 01 ALFA-ROMEO | 26 MERCEDES |
| 02 AUDI | 27 MITSUBISHI |
| 03 AVA | 28 NISSAN |
| 04 BMW | 29 NISSAN |
| 05 CHEVROLET | 30 OLIO |
| 06 CHRYSLER | 31 OPEL |
| 07 CITROËN | 32 PEUGEOT |
| 08 DACIA | 33 PORSCHE |
| 09 DAEWOO | 34 PRADA |
| 10 DAF | 35 RENAULT |
| 11 DODGE | 36 ROVER |
| 12 FIAT | 37 SAAB |
| 13 FORD | 38 SEAT |
| 14 GAZ VOLHA | 39 ŠKODA |
| 15 HOLDEN | 40 STEYR-DANGLER-PUCH |
| 16 HONDA | 41 SUBARU |
| 17 HYUNDAI | 42 SUZUKI |
| 18 IFA | 43 TATA |
| 19 NISSAN | 44 TOYOTA |
| 20 JAGUAR | 45 TRAWANT |
| 21 JEP | 46 VAZ |
| 22 LANCIA | 47 VOLKSWAGEN |
| 23 LAND ROVER | 48 VOLVO |
| 24 LIAZ | 49 WARTBURG |
| 25 MAZDA | 50 ZASTAVA |

79 Jiná výrobní značka osobního automobilu vyrobeného v ČR

60 Jiná výrobní značka osobního automobilu vyrobeného mimo ČR

65 Jiná výrobní značka nákladního automobilu vyrobeného v ČR

66 Jiná výrobní značka nákladního automobilu vyrobeného mimo ČR

68 autobus vyrobený v ČR

69 autobus vyrobený mimo ČR

90 motopád vyrobený v ČR

91 motopád vyrobený mimo ČR

92 motocykl (včetně malého motocyklu) vyrobený v ČR

93 motocykl (včetně malého motocyklu) vyrobený mimo ČR

99 žádná z uvedených (p. pol. 44) je kód 99 až 18)

46b ÚDAJE O VOZIDLE

- u této položky se uvádí:
 - u motocyklů a osobních automobilů zvláštní objem silnic v letech
 - u nákladních automobilů celková hmotnost v letech
 - u osobních automobilů počet míst k sezení
 - u ostatních druhů vozidel se vyplňuje:
 - název (pohlí, například)

48 STÁTNÍ POZNÁVACÍ ZNAČKA

- u vozidel registrovaných v ČR - SPZ
- u vozidel registrovaných mimo území ČR - NPZ

47 ROK VÝROBY VOZIDLA

- poslední dvojmístný rok výroby vozidla

48a CHARAKTERISTIKA VOZIDLA (název vozidla)

- 01 soukromé, nevyjádřené k výdělečné činnosti
- 02 soukromé, využívané k výdělečné činnosti
- 03 soukromá organizace (podnikatel, s.r.o. atd.)
- 04 veřejná hromadná doprava
- 05 nákladní hromadná doprava
- 06 mezinárodní kabinová doprava
- 07 TAXI
- 08 státní podnik, státní organizace
- 09 majetková mimo území ČR
- 10 zastupitelství úřad
- 11 ministerstvo vnitra
- 12 policie ČR
- 13 městská, obecní policie
- 14 soukromé bezpečnostní agentury
- 15 ministerstvo obrany
- 16 jiné
- 17 neobjasněno
- 99 neobjasněno

49b DOPLNĚJÍCÍ ÚDAJE O VOZIDLE

- 1 přeprava nebezpečných nákladů - pevných
- 2 přeprava nebezpečných nákladů - kapalných
- 3 přeprava nebezpečných nákladů - plynových
- 4 přeprava nebezpečných nákladů
- 5 želez s zvláštním výhledovým zrcadlem (zraněný)
- 6 želez s přílevem přednosti (návěs)
- 9 neobjasněno v úvodu

49 SMYK

- 1 ano
- 0 ne

50a VOZIDLO PO NEHODĚ

- 1 nedobro k průběhu
- 2 došlo k poškození
- 3 kód upl - zbytek
- 4 kód upl - neobjasněno, ale vozidlo zůstalo
- 9 žádná z uvedených

50b ÚNIK PROVOZNÍCH, PŘEPRAVOVÝCH HMOT

- 1 došlo k úniku provozních hmot, olej, chladící média apod., z vlastního vozidla
- 2 došlo k úniku jiných nebezpečných látek - pevných
- 3 došlo k úniku jiných nebezpeč. látek - kapalných
- 4 došlo k úniku jiných nebezpeč. látek - plynových
- 9 žádná z uvedených

51 ZPŮSOB VYPROSTĚNÍ OSOB Z VOZIDLA

- 1 nebylo třeba užít násilí
- 2 použitím páčidel apod.
- 3 použitím speciální vyprošťovací techniky

52 SMĚR JÍZDY NEBO POSTAVENÍ VOZIDLA

- 01 jedouc - ve směru stání na komunikaci
- 02 odstavené - ve směru stání na komunikaci
- 03 odzouč - proti směru stání na komunikaci
- 04 odstavené - proti směru stání na komunikaci
- 05 vozidlo jedouc - na komunikaci bez stání na ulici
- 06 vozidlo odstavené, parkující - na komunikaci bez stání na ulici
- 10 - 99 zachycuje postavení vozidla při nehodě na křižovatce

53 ŠKODA NA VOZIDLE

ve stokrátích - vyplí zprava

54 RODNÉ ČÍSLO ŘIDIČE

pouze u řidičů ČR, u ostatních se uvádě datum narození a za lomítkem písmeno C

55a KATEGORIE ŘIDIČE (uvězte nejvyšší skupinu)

- 1 a řidičským oprávněním skupiny A
- 2 a řidičským oprávněním skupiny B
- 3 a řidičským oprávněním skupiny C
- 4 a řidičským oprávněním skupiny D
- 5 a řidičským oprávněním skupiny T
- 6 a řidičským oprávněním skupiny A - do 50 cm³
- 7 bez příslušného řidičského oprávnění
- 8 ostatní řidiči, včetně (vyhlá. voška apod.)
- 9 neobjasněno, kód místo nebylo opatřeno
- 0 neobjasněno (př. u ostatních)

56b NEJVYŠÍ UČENÉ VZDĚLÁNÍ

- 1 základní škola
- 2 učňovská škola
- 3 střední škola
- 4 vysoká škola
- 5 neobjasněno

59 DÉLKA ŘIČIČSKÉ PRAXE V ŘEZIMĚ MOTOROVÉHO VOZIDLA

(v letech, s příslušným druhem vozidla)

67 STAV ŘIDIČE

- 1 dobrý - žádná nepříznivá okolnost nebyla zjištěna
- 2 usazen, uvaž
- 3 pod vlivem léků, nárkotik
- 4 pod vlivem alkoholu
- 5 náhle fyzická nezpůsobilost
- 6 nemoc, ústa apod.
- 7 invalidita
- 8 kód při jízdě zemřel (inaktiv apod.)
- 9 pouze u silničních, sebeznaněno
- 0 jiný nepříznivý stav

68 VNĚJŠÍ OVLIVNĚNÍ ŘIDIČE

- 1 kód nebyl ovlivněn
- 2 oslbné slunce
- 3 oslbné světlo (jiného vozidla)
- 4 ovlivněn jedním z výše uvedených sl. prvků
- 5 ovlivněn při vyhlá. vyhl. domácímu zvláštní
- 6 jiný ovlivněn

50 NÁSLEDKY VE VOZIDLE

- a) omezení osob
 - 1 řidič
 - 2 spolujezdiči na předním sedadle vedle řidiče nebo cestující na motocyklu, případně kolo
 - 3 sezení cestujících na zadním sedadle
 - 4 ostatní spolucestující
- b) úroveň omezení osob
 - 1 s příbiv (pouze u motocyklů, vyhlá. vyhl.)
 - 2 bez příbiv (pouze u motocyklů, vyhlá. vyhl.)
 - 3 příbiv na bezpečnostní pásy (i na zadních sedadlech)
 - 4 nepřipoutané bezpečnostní pásy
 - 5 sezení v druhé řadě
 - 6 vozidlo nevybaveno dle zák. předp. (např. v č. 136)
 - 7 bezpečnostní pás (jez. bag.) v č. 136
- c) počet osob
 - 1 muž
 - 2 žena
 - 3 chlapec (do 15 let)
 - 4 dívka (do 15 let)
- 0 NA nemozem (poslední dvojmístný rok)
- a) státní příslušnost (stát)
- 6 poskytnutí první pomoci
 - 1 nebylo třeba poskytnout
 - 2 poskytnuto osobou vozidla zúčastněných na nehodě
 - 3 jinou osobou
 - 4 lékařskou záchrannou službou
 - 5 vozidlem RZP
 - 6 nebylo poskytnuto, ale bylo nutno poskytnout
- 0) následky
 - 1 úmrtí
 - 2 těžká zranění
 - 3 lehká zranění
 - 4 bez zranění

01 IDENTIFIKAČNÍ ČÍSLO

zpracovatele nehody a pořadové číslo

02 ČASOVÉ ÚDAJE O DOPRAVNÍ NEHODĚ**03 DATUM NAKLÁDEJÍ NEHODY**

počet a datační náležitosti (za 12 a více letů)

04 ÚZEMNÍ MÍSTO DOPRAVNÍ NEHODY

kráj, okres, území místa nehody

05a LOKALITA NEHODY

1 v obci (vyplní se i pod 05b)
2 mimo obec

05b KÓD OBCE (stejný kód obce)**06 DRUH NEHODY**

- 1 srážka s jeholovým nebo ledovým vozidlem
- 2 srážka s vozidlem zaparkovaným, odstaveným
- 3 srážka s pevnou překážkou
- 4 srážka s chodcem
- 5 srážka s ležící zvířetím
- 6 srážka s domácním zvířetem
- 7 srážka s vlakem
- 8 srážka s tramvají
- 9 havárie
- 0 jiný druh nehody

07 DRUH SRAŽKY JEDOUČÍCH VOZIDEL

- 1 čelní
- 2 boční
- 3 z boku
- 4 záseď
- 0 nepřízná (v závěru, nejm. o srážku jedoucích voz.)

08 DRUH PŮVODNÍ PŘEKÁŽKY

- 1 strom
- 2 sloup - telefonní, vod., světelný, at. vedení apod.
- 3 odnětá, jehol., skopce, dopr. značky apod.
- 4 svodidla
- 5 překážka vzniklá průvazem jiného vozidla
- 6 voz. pevná část vozidla, podvozku, tunel apod.
- 7 zábrany železničního přejezdu
- 8 překážka vzniklá stavbou, činností (přenos, dopr. značky, tramvaje, železn., železn. apod.)
- 9 jiné překážky (zábrany, opěrná, násep, relativní sádkové apod.)
- 0 nepřízná (v závěru, nejm. o srážku s pev. překážkou)

09 CHARAKTER NEHODY

- 1 nehoda s následky na životě nebo ztrát
- 2 nehoda pouze s hmotnou škodou

10 ZAVNĚNÍ NEHODY

- 1 řidičem motorového vozidla
- 2 řidičem nenmotorového vozidla
- 3 chodcem
- 4 ležící zvířetím, domácním zvířetem
- 5 jiným účastníkem silničního provozu
- 6 chycenou komunikací
- 7 technickou závadou vozidla
- 0 jiné závěry

11 ALKOHOL U VNĚKA NEHODY PŘÍTOMNÝ

- 1 ano
- 2 ne
- 0 nepřízná

12 HLAVNÍ PŘÍČINY NEHODY

100 nesprávné řízení

NEPŘÍMÝCH RYCHLOST JÍZDY

- 201 nepřiměřeně rychlostí vjezd
- 202 nepřiměřeně rychlostí vjíždění (vjezd, vjezd, vjezd, vjezd na červené světlo apod.)
- 203 nepř. rychlostí vjezdem vozidla a nákladu
- 204 nepř. rychlostí stavu vozovky (nájezd, výhledy, blížn., moklý povrch apod.)
- 205 nepř. rychlostí dopravně technického stavu vozovky (zářezky, kámen, stoupání, šláse apod.)
- 206 přetvoření předepsané rychlostní tabulky
- 207 přetvoření rychlosti stanovené dopravní značkou
- 208 nepř. rychlostí bočního, násovkového vjezdu (př. měř. jízdy, předjíždění vozidel)
- 209 jiný druh nepřiměřené rychlosti

NEPŘÍMÝCH PŘEDJÍŽDĚNÍ

- 301 předjíždění vjezdu
- 302 předjíždění bez dostatečného bočního odstavu
- 303 předjíždění bez dostatečného vzhledu (v nepřiměřené vzdálenosti nebo její blízkosti, před vjezdem stoupání apod.)
- 304 při před. došlo k ohrožení protijed. řidiče (sporný vzhled vozidelnosti k předjíždění apod.)
- 305 při předjíždění došlo k ohrožení předjížděného řidiče (jeholové zastavení, předjížděný řidič prudce brzdil, nájezd směrem jízdy apod.)
- 306 předjíždění vjezdu vozidla odbočujícího vjezdu
- 307 před. v nájezdu, kde je zakázáno dopr. značkou
- 308 při předjíždění přejezdu protějšího směru
- 309 bránění v předjíždění
- 310 překážkou při předjíždění související jedoucím vozidlem
- 311 jiný druh nepřiměřeného předjíždění

NEPŘÍMÝCH PŘEDNOSTI V JÍZDĚ

- 401 jízda na „červené světlo“ (červeného semaforu)
- 402 před. příkaz dopravní značky STUJ. DEJ. PŘEDNOST
- 403 před. příkaz dopravní značky DEJ. PŘEDNOST
- 404 vozidla příjezdící zprava
- 405 při odbočování vlevo
- 406 nemat. ležící odbočuje
- 407 protilevou vozidlo při objížďce překážky
- 408 při zaházení do proudu jedoucích vozidel ze strany, nast. zastavení nebo stání
- 409 při vjíždění na silnici
- 410 při odjezdu nebo odbočování
- 411 při předjíždění z jednoho pruhu do druhého
- 412 chodci na vyznačeném přejezdu
- 413 při odbočování vlevo související jedoucím vozidlem
- 414 jiné než před. přednost

NEPŘÍMÝCH ZPŮSOB JÍZDY

- 501 jízda po nesprávné straně, vjezd do protějšku
- 502 vjíždění bez dostatečné brzdy vjezdu
- 503 neodpovídající nesprávné vyznačení ze vozidel
- 504 nesprávné odjezdy nebo odbočení
- 505 chyby při ujetí směrem jízdy
- 506 bezohledná, agresivní, neochotná jízda
- 507 náhlé bezohledné změny rychlosti jízdy
- 508 řízení se při nevhodném řízení vozidla
- 509 samovolné rozjetí nezastaveného vozidla
- 510 újeť na nevhodném místě
- 511 nezvládnutí řízení vozidla
- 512 jízda (jízdy) jedním směrem ulicí, silnicí v protějšku
- 513 srážka v odbočování (pouč. před. před. k ná. silnicím zastavení (zastavení jízdy atd.)
- 514 nehoda v odbočování (pouč. silnicí (ztrata)
- 515 nehoda při provádění silničního zábrany
- 516 jiný druh nepřiměřeného způsobu jízdy

TECHNICKÁ ZÁVADA VOZIDLA

- 601 závada řízení
- 602 závada provozní brzd
- 603 neúčinná nebo nefungující parkovací brzda
- 604 opoždění brzdoucího působení pod stanovenou mez
- 605 defekty pneumatiky způsobující přizpůsobení nebo ná. jiným úhlem vozidla
- 606 závada související se stavem vozidla (neúčinná, chybějící, znečištěná apod.)
- 607 nepřiměřeně nebo polokázaně spojovací hadice
- 608 poruchy řízení vozidla
- 609 nesprávné umístění nákladu
- 610 špatná, ztráta křídla vozidla (řezaného)
- 611 zablokování kol v odbočování (mechanické závady (za. šňový motor, pneumatické, apod.) řízení apod.)
- 612 lom závěsu kola, pružiny
- 613 nevhodně, polokázaně bočnice (řezaného)
- 614 závada odbočování pro příjezd
- 615 údržba spojovací hadice
- 616 jiné technické závady

13 NÁSLEDKY NEHODY - stav do 24 hod.

- a) zraněná osoba
- b) úmrtí zraněná osoba
- c) ztráta zraněná osoba

14 CELKOVÁ HMOTNÁ ŠKODA

ve stotkách vypl. zprava

15 DRUH PŮVODNÍ VOZIDLOVY

- 1 srážka
- 2 žilva
- 3 beton
- 4 pánev
- 5 bláto
- 6 jiný nesprávný povrch
- 0 žádný z uvedených (bláto, šláse atd.)

16 STAV PŮVODNÍ VOZIDLOVY V DOBĚ NEHODY

- 1 povrch suchý, neznečištěný
- 2 povrch suchý, znečištěný (bláto, šláse, atd.)
- 3 povrch moklý
- 4 na vozovce je bláto
- 5 na vozovce je nájezd, újeť směrem - nepopadl
- 6 na vozovce je nájezd, újeť směrem - nepopadl
- 7 na vozovce je nájezd, újeť směrem - nepopadl
- 8 souvislé odměny vjezdu, rozložení směrem
- 9 náhlé změny stavu vozovky (nájezd na maslu, mláti, nájezd apod.)
- 0 jiný stav povrchu vozovky s dob. nehody

17 STAV KOMUNIKACE

- 01 dobrý, bez závad
- 02 poškození silnicí vypl. nad 5%
- 03 nesprávné umístění, znečištění, chybějící dopravní značky
- 04 zvláštní povrch v podélném směru
- 05 souvislé výhledy
- 06 nesouvislé výhledy
- 07 jiné závady vozovky
- 08 přehradní stěna, hráz, vstoupání, propadlé koleje
- 09 nepřehraděná nebo nedostatečně označená překážka na komunikaci
- 10 přehradní stěna jednoho směru
- 11 přehradní stěna komunikace nebo jiného jezu
- 12 jiný (přehradní) stav nebo závada komunikace

18 POVĚTRNOSTNÍ PODMÍNKY V DOBĚ NEHODY

- 1 nečistota
- 2 mlha
- 3 na počátku deště, sloný déšť
- 4 déšť
- 5 sněžení
- 6 sníh se námrazou, nájezd
- 7 násovkový vjezd (boční, vjezd apod.)
- 8 jiné závady

19 VÍDELNOST

- 1 ve dne, viditelnost nechráněná vjezdem podmíněných podmínek
- 2 ve dne, zhoršená viditelnost (vláha, soumrak)
- 3 ve dne, zhoršená viditelnost vjezdem podmíněných podmínek (mlha, snížení, déšť apod.)
- 4 v noci - s velkým osvětlením, viditelnost nechráněná vjezdem podmíněných podmínek
- 5 v noci - s velkým osvětlením, zhoršená viditelnost vjezdem podmíněných podmínek (mlha, snížení, déšť apod.)
- 6 v noci - bez viditelného osvětlení, viditelnost nechráněná vjezdem podmíněných podmínek
- 7 v noci - bez viditelného osvětlení, viditelnost zhoršená vjezdem podmíněných podmínek (mlha, snížení, déšť apod.)

20 ROZHLEDOVÉ POMĚRY

- 1 dobrý
- 2 špatný vjezdem okrajů silnicí (bujavy, jiné zábrany, ležící apod.)
- 3 špatný vjezdem profilu komunikace (nepřehraděný vjezd, stoupání, zářez komunikace apod.)
- 4 špatný vjezdem trávník vegetace (tráva, křovina apod.)
- 5 špatný vjezdem přechodná vegetace (tráva, tráva)
- 6 vjezd zahný stojícím vozidlem
- 0 jiné špatné

21 ČLENĚNÍ KOMUNIKACE

- 1 dvupruhová
- 2 třípruhová
- 3 čtyřpruhová s odboč. pásem
- 4 čtyřpruhová s odboč. pásem
- 5 vjezdová
- 6 žádná z uvedených

22 SITUOVÁNÍ NEHODY NA KOMUNIKACI

- 1 na jednom pruhu
- 2 na odbočném pruhu
- 3 na krajici
- 4 na odbočném, příjezdovém pruhu
- 5 na pruhu pro pomalá vozidla
- 6 na odbočném nebo odbočce
- 7 na kolejišti tramvaje
- 8 vjezdová komunikace
- 9 na silnici pro cyklisty
- 0 žádná z uvedených

23 ŘÍZENÍ PROVOZU V DOBĚ NEHODY

- 1 policajtem nebo jiným pověřeným orgánem
- 2 zvláštním signačním zařízením
- 3 místní úprava (vyplní se pod 24)
- 0 žádný způsob řízení provozu

24 MÍSTNÍ ÚPRAVA PŘEDNOSTI V JÍZDĚ

- 1 světelná signalizace, průsvětelná žilva
- 2 světelná signalizace mimo provoz
- 3 světelná signalizace dopravními značkami
- 4 světelná signalizace dopravními značkami nebo zařízeními
- 5 světelná signalizace - vypíná se zpravidla
- 0 žádná místní úprava

27 SPECIFICKÁ MÍSTA A OBJEKTY V MÍSTĚ NEHODY

- 01 překážka pro chodce
- 02 v blízkosti přejezdu pro chodce (do 20 m)
- 03 železniční přejezd zabezpečený
- 04 železniční přejezd zabezpečený
- 05 most, nájezd, podjezd, tunel
- 06 zatáčka zatáčková, tramvaje atd. a nájezd, zatáčková
- 07 zatáčka zatáčková, zatáčková atd. bez nájezdu, zatáčková
- 08 vjezd do parkoviště, ležící směrem apod. (pod 26 + 7. B)
- 09 dopadlé pohyblivé vozidlo
- 10 nepřehraděná přehradící se komunikace
- 09 žádná nebo žádná z uvedených

28 SMĚROVÉ POMĚRY

- 1 příčný úsek
- 2 příčný úsek pro směr zatáčkou (do 200 m od místa nehody)
- 3 zatáčka
- 4 zatáčka zatáčková - čtyřpruhová
- 5 zatáčka zatáčková - třipruhová
- 6 zatáčka zatáčková - vjezdová
- 7 vjezdový újezd

29 KATEGORIE VOZIDEL

- 1 nájezd
- 2 žilva
- 3 silnice (do 10 km)
- 4 skupina silnic
- 5 jiné silnice (černá, křiž. skupina atd.)

Příloha B: Datový slovník

Atribut	Datový typ	Popis	Hodnoty
X	Range	Vyjadřuje x-ovou souřadnici GPS	<621202,5;676422,3>
Y	Range	Vyjadřuje y-ovou souřadnici GPS	<1045818;1088912>
p1	Nominal	jedná se o identifikační kód, který je složený z kódu příslušného kraje okresu a útvaru, který nehodu šetřil	unikátní pro každý záznam
p36	Nominal	Druh pozemní komunikace	0,1,2,3,4,5,6,8
p37	Nominal	Číslo pozemní komunikace	unikátní hodnota pro každou komunikaci
p38	Range	Kilometr nehody	<0;433,2>
p2a	Nominal	Datum uskutečnění nehody	<1.1.2006;31.12.2011>
den	Nominal	Den uskutečnění nehody v rámci týdne	0,1,2,3,4,5,6
čas	Range	Přibližný čas uskutečnění nehody	<00:00;23:59>
p6	Nominal	Druh dopravní nehody	1,2,3,4,5,6,7,8,9,0
p7	Nominal	Druh srážky jedoucích vozidel	1,2,3,4,0
p8	Nominal	Druh pevné překážky	1,2,3,4,5,6,7,8,9,0
p9	Nominal	Charakter nehody	1 , 2
p10	Nominal	Zavinění nehody	1,2,3,4,5,6,7,0
p11	Nominal	Přítomnost alkoholu u řidiče	1,2,3,4,5,6,7,8,9,0
p12	Nominal	Hlavní příčiny nehody	100,201-209,301-311,401-414,501-616,601-615
p13a	Range	Usmrceno na následky nehody do 24 hodin	<0;2>
p13b	Range	Těžce zraněno na následky nehody do 24 hodin	<0;4>
p13c	Range	Lehce zraněno na následky nehody do 24 hodin	<0;11>
p14	Range	Celková hmotná škoda (ve 100 Kč)	<0;47400>
p15	Nominal	Druh povrchu vozovky	1,2,3,4,5,6,0
p16	Nominal	Stav povrchu v době nehody	1,2,3,4,5,6,7,8,9,0
p17	Nominal	Stav komunikace	01,02,03,04,05,06,07,08,09,10,11,12
p18	Nominal	Povětrnostní podmínky v době nehody	1,2,3,4,5,6,7,0
p19	Nominal	Viditelnost v době nehody	1,2,3,4,5,6,7
p20	Nominal	Rozhledové poměry	1,2,3,4,5,6,0
p21	Nominal	Dělení komunikace	1,2,3,4,5,6,0
p22	Nominal	Situování nehody na komunikaci	1,2,3,4,5,6,7,8,9,0
p23	Nominal	Řízení provozu v době nehody	1,2,3,0
p24	Nominal	Místní úprava přednosti v jízdě	1,2,3,4,5,0
p27	Nominal	Specifická místa a objekty v době nehody	01,02,03,04,05,06,07,08,09,10,00

p28	Nominal	Směrové poměry	1,2,3,4,5,6,7
p34	Range	Počet zúčastněných vozidel	<1;10>
p35	Nominal	Místo dopravní nehody	00,10,11-18,19,22-28,29
p39	Nominal	Druh křižující komunikace	1,2,3,6,7,9
p44	Nominal	Druh vozidla	00,01,02,03,04,05,06,07,08,09,10,11,12,13,14,15,16,17,18
p45a	Nominal	Výrobní značka motorového vozidla	01,02,03,04,05,06,07,08,09,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,00
p47	Nominal	Rok výroby vozidla	<1954;2011>
p48a	Nominal	Charakter vozidla	01,02,03,04,05,06,07,08,09,10,11,12,13,14,15,16,17,18,00
p49	Flag	Důvodem nehody byl smyk	0, 1
p50a	Nominal	Stav vozidla po nehodě	1,2,3,4,0
p50b	Nominal	Únik provozních či přepravujících látek	1,2,3,4,0
p51	Nominal	Způsob vyproštění osob z vozidla	1,2,3
p52	Nominal	Směr jízdy nebo postavení vozidla	01,02,03,04,05,06,10-99
p53	Range	Škoda na vozidle (ve 100 Kč)	<0;20000>
p55a	Nominal	Kategorie řidiče	1,2,3,4,5,6,7,8,9,0
p57	Nominal	Stav řidiče	1,2,3,4,5,6,7,8,9,0
p58	Nominal	Vnější ovlivnění řidiče	1,2,3,4,5,0

Zdroj: [dle dat]

Příloha C: Charakteristiky kategoriálních atributů.

Hodnota	p36		den		p6		p7		p8	
	počet	poměr	počet	poměr	počet	poměr	počet	poměr	počet	poměr
0	117	0.86%	1404	10.31%	336	2.47%	6082	44.66%	11670	85.70%
1	4226	31.03%	1340	9.84%	7529	55.29%	514	3.77%	381	2.80%
2	2254	16.55%	2230	16.38%	1973	14.49%	1830	13.44%	284	2.09%
3	1922	14.11%	2041	14.99%	1948	14.30%	2097	15.40%	245	1.80%
4	-	-	2091	15.35%	324	2.38%	3095	22.73%	156	1.15%
5	-	-	2158	15.85%	666	4.89%	-	-	20	0.15%
6	4297	31.55%	2354	17.29%	55	0.40%	-	-	251	1.84%
7	61	0.45%	-	-	7	0.05%	-	-	27	0.20%
8	741	5.44%	-	-	-	-	-	-	31	0.23%
9	-	-	-	-	780	5.73%	-	-	553	4.06%
Hodnota	p9		p10		p11		p15		p16	
	počet	poměr	počet	poměr	počet	poměr	počet	poměr	počet	poměr
0	-	-	60	0.44%	3181	23.36%	29	0.21%	22	0.16%
1	2646	19.43%	12114	88.96%	437	3.21%	557	4.09%	9720	71.38%
2	10972	80.57%	527	3.87%	9874	72.51%	12620	92.67%	58	0.43%
3	-	-	94	0.69%	53	0.39%	296	2.17%	2947	21.64%
4	-	-	718	5.27%	2	0.01%	41	0.30%	12	0.09%
5	-	-	18	0.13%	-	-	48	0.35%	204	1.50%
6	-	-	44	0.32%	5	0.04%	27	0.20%	482	3.54%
7	-	-	43	0.32%	3	0.02%	-	-	7	0.05%
8	-	-	-	-	17	0.12%	-	-	161	1.18%
9	-	-	-	-	46	0.34%	-	-	5	0.04%
Hodnota	p18		p19		p20		p21		p22	
	počet	poměr	počet	poměr	počet	poměr	počet	poměr	počet	poměr
0	17	0.12%	-	-	11	0.08%	714	5.24%	263	1.93%
1	12077	88.68%	10052	73.81%	13555	99.54%	11799	86.64%	12348	90.67%
2	99	0.73%	215	1.58%	11	0.08%	295	2.17%	33	0.24%
3	296	2.17%	462	3.39%	19	0.14%	454	3.33%	298	2.19%
4	718	5.27%	1392	10.22%	6	0.04%	245	1.80%	13	0.10%
5	291	2.14%	159	1.17%	0	0.00%	103	0.76%	5	0.04%
6	105	0.77%	1189	8.73%	16	0.12%	8	0.06%	76	0.56%
7	15	0.11%	149	1.09%	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	527	3.87%
9	-	-	-	-	-	-	-	-	55	0.40%

Hodnota	p23		p24		p28		p39		p49	
	počet	poměr	počet	poměr	počet	poměr	počet	poměr	počet	poměr
0	9584	70.38%	10035	73.69%	-	-	-	-	11087	92.89%
1	19	0.14%	-	-	8247	60.56%	239	6.22%	848	7.11%
2	456	3.35%	-	-	535	3.93%	282	7.34%	-	-
3	3559	26.13%	3048	22.38%	994	7.30%	709	18.45%	-	-
4	-	-	241	1.77%	1637	12.02%	-	-	-	-
5	-	-	294	2.16%	2002	14.70%	-	-	-	-
6	-	-	-	-	20	0.15%	2388	62.16%	-	-
7	-	-	-	-	183	1.34%	218	5.67%	-	-
8	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	6	0.16%	-	-
Hodnota	p50a		p50b		p51		p55a		p57	
	počet	poměr	počet	poměr	počet	poměr	počet	poměr	počet	poměr
0	664	5.56%	11563	96.88%	-	-	21	0.18%	17	0.15%
1	11053	92.61%	372	3.12%	11854	99.32%	29	0.24%	11042	94.47%
2	15	0.13%	-	-	27	0.23%	7262	60.91%	44	0.38%
3	85	0.71%	-	-	54	0.45%	3140	26.34%	12	0.10%
4	118	0.99%	-	-	-	-	605	5.07%	441	3.77%
5	-	-	-	-	-	-	4	0.03%	120	1.03%
6	-	-	-	-	-	-	4	0.03%	2	0.02%
7	-	-	-	-	-	-	116	0.97%	2	0.02%
8	-	-	-	-	-	-	507	4.25%	7	0.06%
9	-	-	-	-	-	-	235	1.97%	1	0.01%
Hodnota	p58									
	počet	poměr								
0	15	0.13%								
1	11616	99.38%								
2	27	0.23%								
3	6	0.05%								
4	9	0.08%								
5	15	0.13%								
6	-	-								
7	-	-								
8	-	-								
9	-	-								

Hodnota	p17		p27		p35		p44		p48a	
	počet	poměr	počet	poměr	počet	poměr	počet	poměr	počet	poměr
00	-	-	10523	77.27%	9776	87.59%	19	0.16%	11	0.09%
01	13301	97.67%	782	5.74%	-	-	21	0.18%	7336	61.48%
02	3	0.02%	861	6.32%	-	-	142	1.19%	266	2.23%
03	10	0.07%	10	0.07%	-	-	8421	70.56%	3681	30.85%
04	16	0.12%	32	0.23%	-	-	107	0.90%	34	0.28%
05	20	0.15%	270	1.98%	-	-	1659	13.90%	87	0.73%
06	56	0.41%	35	0.26%	-	-	210	1.76%	5	0.04%
07	4	0.03%	107	0.79%	-	-	577	4.83%	10	0.08%
08	6	0.04%	71	0.52%	-	-	146	1.22%	46	0.39%
09	2	0.01%	95	0.70%	-	-	38	0.32%	309	2.59%
10	5	0.04%	832	6.11%	1342	12.02%	0	0.00%	0	0.00%
11	4	0.03%	-	-	14	0.13%	54	0.45%	23	0.19%
12	191	1.40%	-	-	10	0.09%	33	0.28%	85	0.71%
13	-	-	-	-	15	0.13%	503	4.21%	5	0.04%
14	-	-	-	-	4	0.04%	3	0.03%	6	0.05%
15	-	-	-	-	-	-	2	0.02%	22	0.18%
16	-	-	-	-	-	-	-	-	3	0.03%
17	-	-	-	-	-	-	-	-	-	-
18	-	-	-	-	-	-	-	-	3	0.03%

Hodnota	p12	
	počet	poměr
100	934	6.86%
201 - 209	2046	15.02%
301 - 311	383	2.81%
401 - 414	2269	16.66%
501 - 516	7943	58.33%
601 - 615	43	0.32%

Hodnota		p47	Hodnota		p47
	počet	poměr		počet	poměr
54	2	0.02%	86	116	1.02%
57	1	0.01%	87	131	1.16%
59	2	0.02%	88	158	1.39%
60	2	0.02%	89	198	1.75%
63	1	0.01%	90	287	2.53%
64	1	0.01%	91	229	2.02%
65	3	0.03%	92	329	2.90%
66	1	0.01%	93	299	2.64%
67	3	0.03%	94	309	2.73%
68	3	0.03%	95	486	4.29%
69	1	0.01%	96	630	5.56%
70	4	0.04%	97	751	6.63%
71	1	0.01%	98	708	6.25%
72	10	0.09%	99	740	6.53%
73	10	0.09%	00	637	5.62%
74	11	0.10%	01	504	4.45%
75	14	0.12%	02	562	4.96%
76	13	0.11%	03	629	5.55%
77	21	0.19%	04	606	5.35%
78	19	0.17%	05	679	5.99%
79	34	0.30%	06	712	6.29%
80	27	0.24%	07	551	4.86%
81	22	0.19%	08	340	3.00%
82	22	0.19%	09	173	1.53%
83	27	0.24%	10	111	0.98%
84	58	0.51%	11	45	0.40%
85	94	0.83%			

Hodnota		p48a	Hodnota		p48a
	počet	poměr		počet	poměr
00	633	5.30%	42	104	0.87%
01	17	0.14%	43	68	0.57%
02	157	1.32%	44	123	1.03%
03	118	0.99%	45	6	0.05%
04	148	1.24%	46	44	0.37%
05	12	0.10%	47	586	4.91%
06	25	0.21%	48	179	1.50%
07	294	2.46%	49	12	0.10%
08	10	0.08%	51	3	0.03%
09	68	0.57%	55	8	0.07%
10	175	1.47%	59	67	0.56%
11	9	0.08%	60	77	0.65%
12	385	3.23%	61	13	0.11%
13	1188	9.95%	62	157	1.32%
16	155	1.30%	64	8	0.07%
17	164	1.37%	65	2	0.02%
18	1	0.01%	68	1	0.01%
19	259	2.17%	70	1	0.01%
21	7	0.06%	71	2	0.02%
22	10	0.08%	73	1	0.01%
23	11	0.09%	78	6	0.05%
24	66	0.55%	79	3	0.03%
25	128	1.07%	82	1	0.01%
26	365	3.06%	83	1	0.01%
27	79	0.66%	84	1	0.01%
29	114	0.96%	85	2	0.02%
31	541	4.53%	86	1	0.01%
32	542	4.54%	87	1	0.01%
33	4	0.03%	88	19	0.16%
34	7	0.06%	89	8	0.07%
35	803	6.73%	93	2	0.02%
36	7	0.06%	94	1	0.01%
37	5	0.04%	95	5	0.04%
38	167	1.40%	97	33	0.28%
39	3434	28.77%	98	1	0.01%
40	121	1.01%	99	149	1.25%
41	10	0.08%			

Příloha D: CD s vytvořenými modely a analyzovanými daty

viz volně vložené CD ve vytištěné bakalářské práci