



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA DOPRAVNÍ

Bc. Martin Zajíček

**VYUŽITÍ BAYESOVSKÝCH SÍTÍ
PRO ANALÝZU NEHODOVÝCH DAT**

Diplomová práce

2023



K620..... **Ústav dopravní telematiky**

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

Jméno a příjmení studenta (včetně titulů):

Bc. Martin Zajíček

Studijní program (obor/specializace) studenta:

navazující magisterský – IS – Inteligentní dopravní systémy

Název tématu (česky): **Využití Baysovských sítí pro analýzu nehodových dat**

Název tématu (anglicky): Use of Bayesian networks for accident data analysis

Zásady pro vypracování

Při zpracování diplomové práce se řiďte následujícími pokyny:

- Rešerše metod analýzy nehodových dat včetně rešerše použití Baysovských sítí pro tento účel;
- Popis a rozbor dostupných datových sad pro použití v rámci diplomové práce;
- Vytvoření modelu nehodovosti pomocí Baysovských sítí na základě naměřených dat;
- Návrh postupu pro procházení marginálních a pravděpodobnostních funkcí pro definované příklady/otázky;
- Porovnání výsledků analýzy nehod prostřednictvím Baysovských sítí s jinými metodami.





- Rozsah grafických prací: dle požadavků vedoucích diplomové práce
- Rozsah průvodní zprávy: minimálně 55 stran textu (včetně obrázků, grafů a tabulek, které jsou součástí průvodní zprávy)
- Seznam odborné literatury: SCUTARI, Marco a Jean-Baptiste DENIS. Bayesian Networks [online]. Boca Raton: Chapman and Hall/CRC, 2021 [cit. 2023-04-18]. ISBN 9780429347436. Dostupné z: doi:10.1201/9780429347436


Vedoucí diplomové práce: **Ing. Jiří Růžička, Ph.D.**
Ing. Miroslav Vaniš, Ph.D.

Datum zadání diplomové práce: **1. června 2022**
(datum prvního zadání této práce, které musí být nejpozději 10 měsíců před datem prvního předpokládaného odevzdání této práce vyplývajícího ze standardní doby studia)

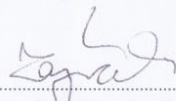
Datum odevzdání diplomové práce: **15. května 2023**
a) datum prvního předpokládaného odevzdání práce vyplývající ze standardní doby studia a z doporučeného časového plánu studia
b) v případě odkladu odevzdání práce následující datum odevzdání práce vyplývající z doporučeného časového plánu studia


Ing. Zuzana Bělinová, Ph.D.
vedoucí
Ústavu dopravní telematiky


prof. Ing. Ondřej Příbyl, Ph.D.
děkan fakulty



Potvrzuji převzetí zadání diplomové práce.


Bc. Martin Zajíček
jméno a podpis studenta

V Praze dne.....1. června 2022

Poděkování

Rád bych na tomto místě poděkoval vedoucím práce Ing. Miroslavu Vanišovi, Ph.D. a Ing. Jiřímu Růžičkovi, Ph.D. za čas, rady a postřehy k práci, které mi poskytli. Další poděkování patří Ing. Bc. Karlu Kociánovi, Ph.D. a Ing. Jakubovi Nováčkovi Ph.D. z Ústavu soudního znalectví v dopravě ČVUT FD za zprostředkování datových souborů a konzultace k jejich zpracování.

V neposlední řadě bych také chtěl poděkovat své rodině a přátelům za veškerou podporu během celého studia.

Prohlášení

Předkládám tímto k posouzení a obhajobě diplomovou práci, zpracovanou na závěr studia na ČVUT v Praze Fakultě dopravní.

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o etické přípravě vysokoškolských závěrečných prací.

Nemám žádný důvod proti užití tohoto školního díla ve smyslu § 60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze, dne 15.5.2023

.....
Bc. Martin Zajíček

Abstrakt

Diplomová práce se zabývá využitím Bayesovských sítí pro analýzu nehodových dat. Data pochází z období let 2011-2021 z území hlavního města Prahy a přilehlých okresů Praha-východ a Praha-západ. Cílem práce je pomocí Bayesovských sítí identifikovat faktory, jejichž pravděpodobnost se nejvíce zvýší při těžkých nehodách, tedy faktory, které se se zvyšující se vážností nehody vyskytují častěji. Faktory byly identifikovány na základě posouzení rozdílů mezi marginálními a podmíněnými pravděpodobnostními funkcemi jednotlivých faktorů (proměnných). V práci je popsána teorie týkající se Bayesovských sítí, dále postup přípravy a zpracování dat, následné vytvoření struktury Bayesovské sítě a výpočet lokálních pravděpodobnostních rozdělení pomocí algoritmu. V závěru práce jsou shrnuty faktory vyskytující se při vyšší vážnosti nehody.

Klíčová slova

Bayesovské sítě, datová analýza, nehodová data, dopravní nehoda, vážnost nehody

Abstract

The thesis deals with the use of Bayesian networks for accident data analysis. The data comes from the period 2011-2021 from the territory of the capital city of Prague and the adjacent districts of Prague-East and Prague-West. The aim of the thesis is to identify the factors whose probability increases most in severe accidents, i.e. the factors that occur more frequently with increasing accident severity, using Bayesian networks. The factors were identified by comparing the differences between the marginal and conditional probability functions of each factor (variable). The thesis describes the theory related to Bayesian networks, the procedure of data preparation and processing, the subsequent construction of the Bayesian network structure and the calculation of local probability distributions using the algorithm. The thesis concludes with a summary of the factors occurring in higher severity accidents.

Keywords

Bayesian networks, data analysis, accident data, traffic accident, accident severity

Obsah

Seznam zkratk	5
1 Úvod	6
2 Teorie	8
2.1 Teorie grafů	8
2.1.1 Graf, uzel, hrana	8
2.1.2 Rodiče, děti, rodina v grafu	10
2.1.3 Předchůdci, následníci	10
2.2 Teorie pravděpodobnosti	11
2.2.1 Náhodný pokus, náhodný jev, pravděpodobnost	11
2.2.2 Náhodná veličina, diskrétní a spojitá náhodná veličina	12
2.2.3 Rozdělení náhodné veličiny	12
2.2.4 Sdružená, marginální, podmíněná pravděpodobnost	14
2.3 Bayesovské sítě	16
2.3.1 Řetězové pravidlo, Bayesův vzorec	16
2.3.2 Marginální nezávislost, podmíněná závislost, podmíněná nezávislost	17
2.3.3 Definice Bayesovské sítě	18
2.3.4 Algoritmus pro tvorbu sítě	19
2.3.5 Příklad Bayesovské sítě a práce s ní	20
2.4 Shrnutí teoretické kapitoly	24
3 Zpracování nehodových dat	25
3.1 Bayesovské sítě a datová analýza	25
3.2 Datový soubor	26
3.2.1 Použitý software	26
3.2.2 Charakteristika zpracovávané oblasti	26
3.2.3 Vývoj nehodovosti	29
3.2.4 Popis datového souboru	31
3.2.5 Vyřazené a sloučené parametry	32
3.2.6 Redukce počtu stavů parametrů	35
3.3 Výstupy	52
3.3.1 Popis vytvořené Bayesovské sítě	52
3.3.2 Popis zjištěných výstupů	55
4 Závěr	59
Seznam zdrojů	61
Seznam příloh	63
Seznam tabulek	64
Seznam obrázků	65
Seznam grafů	66

Seznam zkratek

CDV	Centrum dopravního výzkumu, v. v. i.
HDP	hrubý domácí produkt
ŘSD ČR	Ředitelství silnic a dálnic České republiky
SSZ	světelné signalizační zařízení
TSK	Technická správa komunikací hl. m. Prahy
vozokm	vozokilometr
VO	veřejné osvětlení

1 Úvod

Dopravní nehody jsou celosvětově velkým problémem v oblasti veřejné bezpečnosti a způsobují značné ztráty na životech, škody na majetku a ekonomické ztráty. Podle Světové zdravotnické organizace zemře každý rok v důsledku dopravních nehod přibližně 1,3 milionu lidí a dalších 20-50 milionů lidí utrpí různě závažná zranění. Dopravní nehody jsou celosvětově nejčastější příčinou úmrtí osob ve věku 5-29 let. Zranění z dopravních nehod způsobují značné ekonomické ztráty jednotlivcům, jejich rodinám i celým státům. Tyto ztráty vznikají v důsledku nákladů na léčbu a ztráty produktivity osob, které byly v důsledku zranění usmrceny nebo hendikepovány, a rodinných příslušníků, kteří si kvůli péči o zraněné musí vzít volno v práci nebo ve škole. Dopravní nehody stojí většinu zemí světa zhruba 3 % jejich HDP. [1]

V České republice došlo podle statistik nehodovosti Policejního prezidia ČR v roce 2022 k celkem 98 460 nehodám, z nichž při každé páté došlo ke zranění nebo usmrcení člověka. Celkem bylo usmrceno 528 osob (454 do 24 hodin do nehody, zbylých 74 do 30 dnů po nehodě) a dalších více než 24 000 jich bylo zraněno, hmotná škoda z těchto nehod se odhaduje na asi 7,5 miliardy Kč, celkové ekonomické ztráty pak na asi 135 miliard Kč. [2] Ačkoliv počty obětí v posledních letech klesají, stále se jedná o vysoké číslo, proto jsou neustále hledány nové způsoby a přístupy, jak přispět ke zvýšení bezpečnosti dopravy.

K dopravním nehodám mohou přispívat různé faktory, včetně lidských faktorů, faktorů vozidla a faktorů životního prostředí. K lidským faktorům patří chování řidičů, jako je překračování rychlosti, řízení pod vlivem alkoholu, rozptýlená pozornost řidičů nebo únava. K faktorům vozidla patří zejména závady vozidla, například vadné brzdy nebo pneumatiky. Faktory prostředí zahrnují stav vozovky, počasí, viditelnost a další faktory prostředí. Pochopení příčin a faktorů, které přispívají k dopravním nehodám, má zásadní význam pro prevenci a snížení jejich četnosti a závažnosti.

Právě k lepšímu pochopení příčin dopravních nehod přispívá i využití datové analýzy, konkrétně Bayesovské sítě. Bayesovské sítě jsou pravděpodobnostní grafický model, který lze použít k modelování složitých systémů a vztahů mezi proměnnými. Jsou zvláště užitečné pro analýzu dat s více proměnnými, mezi kterými existuje může existovat (někdy skrytá) kauzalita, jako jsou právě údaje o dopravních nehodách. Bayesovské sítě jsou také flexibilnější než tradiční statistické metody, jako je regresní analýza, protože si dokážou poradit s nelineárními vztahy a složitými interakcemi mezi proměnnými. Mezi další oblasti, ve kterých lze Bayesovské sítě využívat, patří mj. diagnostika a predikce v medicíně, rozpoznávání řeči a zpracování přirozeného jazyka nebo bezpečnost a kybernetická bezpečnost.

Tato práce se zabývá využitím Bayesovských sítí při analýze nehodových dat. Jedná se o nehody z období let 2011-2021 z území hlavního města Prahy a přilehlých okresů Praha-východ a Praha-západ. Zdrojem dat je Policejní prezidium ČR. Cílem práce je identifikovat stavy jednotlivých parametrů, jejichž

pravděpodobnost se nejvíce zvýší při těžkých nehodách, tedy stavy, které se se zvyšující se vážností nehody vyskytují častěji. Toho lze dosáhnout právě pomocí Bayesovských sítí.

V úvodu práce je představena teorie grafů a teorie pravděpodobnosti, které jsou nezbytné pro následnou definici Bayesovských sítí. Dále jsou popsány způsoby tvorby struktury sítě a na ukázkovém příkladu je vysvětlen princip fungování Bayesovských sítí a jevy s nimi spojené. Další kapitola se zabývá charakteristikou zpracovávaných oblastí a shrnuje základní údaje týkající se dopravních nehod v daných územích. Je popsán datový soubor a postup jeho zpracování, včetně popisu zaznamenávaných parametrů a četnosti jednotlivých hodnot. Vysvětluje také, které parametry byly vynechány, které sloučeny s jinými parametry a u kterých došlo k redukci stavů. Po přípravě dat je popsána tvorba sítě a okomentována její struktura z hlediska uzlů a vazeb mezi nimi. Závěr práce je věnován analýze faktorů, které se nejčastěji vyskytují při vyšší vážnosti nehody a doporučením, na které části je možné se dále v analýze zaměřit, případně další směry, kde lze využít Bayesovské sítě.

2 Teorie

V této kapitole je popsán teoretický aparát, ze kterého vychází praktická část práce. Jednotlivé části se týkají popisu oblastí a základních principů, na kterých fungují Bayesovské sítě. Dvě základní oblasti, ze kterých čerpají a které jsou nezbytné k pochopení principů jejich fungování, jsou teorie grafů a teorie pravděpodobnosti. Každé z nich je věnována jedna podkapitola, třetí podkapitolu tvoří teorie týkající se samotných Bayesovských sítí.

2.1 Teorie grafů

Teorie grafů je jednou z mladších disciplín matematiky. Dnes našla teorie grafů jako významná část diskrétní matematiky své uplatnění při řešení mnoha praktických úloh. Její výhodou je snadná a intuitivní představa, která modeluje reálnou situaci jako množinu objektů (vrcholy grafu) a vztahy mezi nimi (hrany grafu). Převědeme-li konkrétní úlohu do řeči teorie grafů, často se jedná o již řešený obecný problém a můžeme použít známé algoritmy při implementaci řešení. Bezesporu zásadní výhodou je snadné počítačové zpracování.

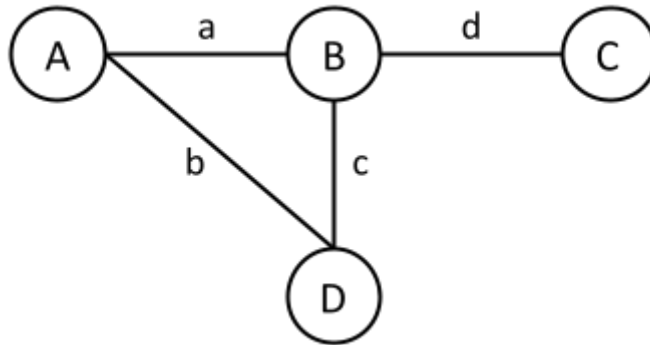
V následujících podkapitolách jsou popsány základní pojmy z teorie grafů, které jsou používány dále v práci. Základním zdrojem, ze kterého vychází popisované pojmy, jsou vysokoškolská skripta Teorie grafů – aplikace v dopravě a veřejné správě. [3]

2.1.1 Graf, uzel, hrana

Základním pojmem v teorii grafů je graf. Graf je uspořádaná trojice $G = (V, X, p)$, kde prvky množiny V jsou nazývány uzly grafu G , prvky množiny X hrany grafu G a p je prosté zobrazení množiny X do množiny všech neuspořádaných dvojic (u, v) , $u, v \in V$, $u \neq v$. Množina V je konečná a neprázdná, tzn. vždy má alespoň jeden prvek. Zobrazení p se také může nazývat incidencí grafu G .

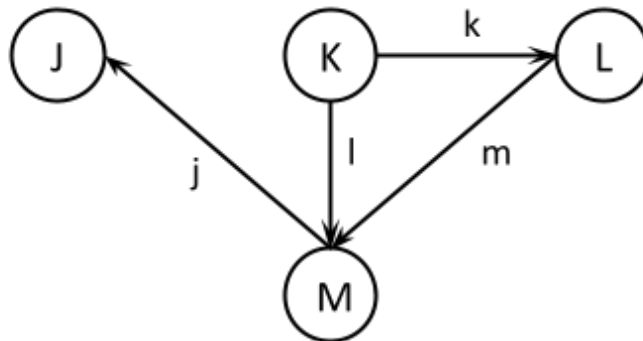
V případě neuspořádaných dvojic (u, v) je za ekvivalent dvojice považována dvojice (v, u) . Graf je potom nazýván neorientovaný. Pokud ale zobrazení p zobrazuje uspořádané dvojice $[u, v]$, $u, v \in V$, $u \neq v$, jedná se o graf orientovaný. Zatímco u neorientovaných hran jsou oba uzly nazývány sousedními, u orientovaných se rozlišují výchozí a koncové uzly. Označení dvojice kulatými závorkami znamená neuspořádanou dvojici, hranaté závorky dvojici uspořádanou. Graficky jsou orientované hrany vyjádřeny šipkou z výchozího do koncového uzlu.

Na obr. 1 je příklad neorientovaného grafu, který tvoří uzly grafu A, B, C, D a neorientované hrany a, b, c, d a zobrazení hran $a = (A, B)$, $b = (A, D)$, $c = (B, D)$, $d = (B, C)$.



obr. 1 - Příklad neorientovaného grafu [zdroj vlastní]

Na obr. 2 je příklad orientovaného grafu tvořeného uzly J, K, L, M a orientovanými hranami j, k, l, m a zobrazení hran $j = [M, J], k = [K, L], l = [K, M], m = [L, M]$.

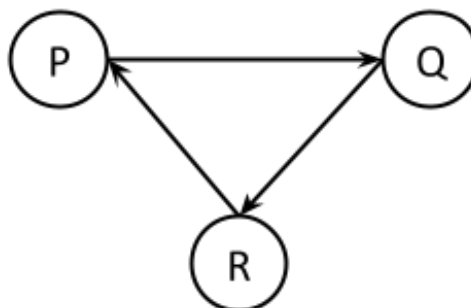


obr. 2 - Příklad orientovaného grafu [zdroj vlastní]

Střídavá posloupnost uzlů a orientovaných hran, ve které se neopakuje žádný z uzlů, se nazývá dráha nebo orientovaná cesta. Uzavřená dráha, kde je výchozí vrchol stejný jako koncový, se nazývá cyklus. Z příkladu na obr. 2 lze definovat dráhu z uzlu K do uzlu J dokonce dvěma způsoby, buď po hranách l, j , nebo po hranách k, m, j .

Graf obsahující alespoň 1 cyklus se nazývá cyklický, naopak graf bez cyklů se nazývá acyklický.

Na obr. 3 je zobrazen příklad cyklu v grafu, příkladem acyklického grafu je graf na obr. 2.

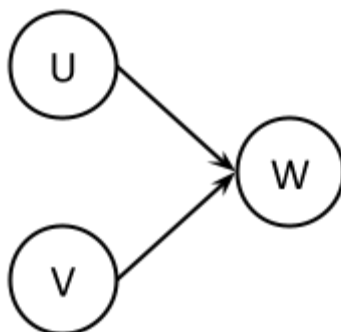


obr. 3 - Příklad cyklu v grafu [zdroj vlastní]

2.1.2 Rodiče, děti, rodina v grafu

Pojmy rodiče, děti a rodina jsou definovány pouze pro orientované grafy, kde je možné určit pořadí uzlů v jednotlivých dráhách. V neorientovaných grafech není možné tyto pojmy vymezit. V případě orientované hrany $[U, V]$ je možné počáteční uzel U označit jako rodiče uzlu V a koncový uzel V jako dítě uzlu U . Pojem rodina uzlu je potom definován jako uzel samotný a všichni jeho rodiče.

Na obr. 4 je graf, kde uzly U, V jsou rodiči uzlu W a naopak uzel W je dítě uzlů U, V . Zatímco rodinu uzlů U, V tvoří vždy pouze samotný uzel, rodinu uzlu W v tomto případě tvoří všechny tři uzly.

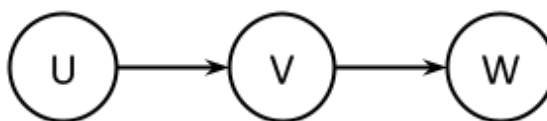


obr. 4 - Příklad rodičů a dětí v grafu [zdroj vlastní]

2.1.3 Předchůdci, následníci

Pojmy předchůdci a následníci se také týkají pouze orientovaných grafů a navazují na pojmy zmíněné v předchozí kapitole.

Pokud je definována orientovaná hrana $[U, V]$, ve které uzel U je rodič uzlu V a uzel V dítě uzlu U , a orientovaná hrana $[V, W]$, ve které uzel V je rodič uzlu W a uzel W dítě uzlu V , jak je zobrazeno na obr. 5, potom lze uzel W označit za následníka uzlu U , a naopak uzel U za předchůdce uzlu W . Zároveň platí, že uzel V je taktéž následníkem uzlu U a že uzel U je předchůdcem uzlu V .



obr. 5 - Příklad předchůdců a následníků v grafu [zdroj vlastní]

2.2 Teorie pravděpodobnosti

Teorie pravděpodobnosti je společně s matematickou statistikou matematická disciplína spadající do vědního oboru, který se nazývá stochastika. Teorie pravděpodobnosti se zabývá popisem zákonitostí týkající se jevů, které (přinejmenším z hlediska pozorovatele) mohou a nemusí nastat, resp. jejichž výsledná hodnota není předem jistá. Tyto jevy se nazývají náhodnými. [4][5]

V následující kapitole budou popsány základní pojmy z teorie pravděpodobnosti, ze kterých vychází Bayesovské sítě.

Jedná se o popis pojmů náhodný jev, náhodný pokus a náhodná veličina, dále rozdělení náhodné veličiny, pravděpodobnostní a distribuční funkce. V závěru kapitoly budou definovány sdružená, marginální a podmíněná pravděpodobnost a jejich vzájemný vztah. Základním zdrojem, ze kterého vychází popis pojmů v této kapitole, jsou vysokoškolská skripta Statistika. [6]

2.2.1 Náhodný pokus, náhodný jev, pravděpodobnost

Náhodným pokusem je myšlen experiment, který za relativně stálých podmínek dává různý výsledek. Důležitý je v tomto případě důraz na slovo relativně, co se týče stálosti podmínek. V reálném světě nelze zaručit absolutní stálost podmínek pro více opakování experimentu, pro potřeby experimentů se ovšem takto uvažuje. Příkladem náhodného pokusu může být hod mincí.

Pro každý náhodný pokus lze definovat tzv. základní prostor, který se skládá z výčtu všech možných výsledků náhodného pokusu. Pro Bayesovské sítě je podstatné, že počet možných výsledků je buď konečný, nebo alespoň popsán spojitou funkcí. Aby byl popis náhodného pokusu kompletní, je nutné ho doplnit o zjištěnou četnost možných výsledků.

Náhodný jev je definován jako libovolný výrok o výsledku náhodného pokusu. Množina všech jevů se nazývá jevové pole. Příkladem náhodného jevu může být výrok „na minci padne orel“. Tento výrok má stejný význam jako výrok „na minci nepadne panna“, jedná se o tzv. množinovou reprezentaci jevu. V případě, že na minci orel opravdu padne, resp. nepadne panna, lze říci, že daný jev nastal.

Posledním pojmem, který doplňuje popis náhodného pokusu, je pravděpodobnost. Pravděpodobnost lze definovat více způsoby, které závisí na pohledu na pojem pravděpodobnosti. Zřejmě nejčastěji se využívá statistická definice, která pravděpodobnost definuje jako podíl počtu experimentů, kdy jev nastal, vůči počtu provedených experimentů.

Matematicky lze tento vztah vyjádřit

$$P(X) = \frac{N_X}{N}$$

kde $P(X)$ je pravděpodobnost jevu X , N_X počet experimentů, kdy jev X nastal, a N celkový počet experimentů. Trojice základní prostor, jevové pole, pravděpodobnost se nazývá pravděpodobnostní prostor a tvoří kompletní pravděpodobnostní popis náhodného pokusu.

2.2.2 Náhodná veličina, diskrétní a spojitá náhodná veličina

Pokud bychom chtěli provádět náhodný pokus v reálném světě, sestavení úplného pravděpodobnostního prostoru by znamenalo komplexní analýzu všech prvků, které pokus zahrnuje, a jejich vlastností, což je velmi často nemožné. Lze ale pracovat s neúplným, nedokonalým popisem – charakteristikou.

Charakteristiky vyžadují numerické výpočty, je tedy nutné výsledkům náhodných pokusů jako rub nebo líc u hodu mincí, zelená nebo červená na SSZ apod. přiřadit číselnou hodnotu. Z výše popsaného vychází definice náhodné veličiny. Náhodná veličina X je definována jako zobrazení z množiny výsledků náhodného pokusu Ω do množiny reálných čísel.

Ve vztahu k náhodnému pokusu lze říci, že náhodná veličina zastupuje náhodný pokus, a hodnoty náhodné veličiny odpovídají výsledkům náhodného pokusu. Hodnoty náhodné veličiny se také nazývají realizace.

Náhodnou veličinu můžeme rozdělit na diskrétní a spojitou podle toho, jak je dána množiny jejich realizací. Pokud je množina realizací konečná nebo spočetná, jde o diskrétní náhodnou veličinu. Příkladem je přítomnost vozidla na detektoru nebo signál na SSZ. Pokud je množina realizací nespočetná, jedná se o spojitou náhodnou veličinu. Příkladem může být změřená rychlost vozidla nebo časová mezera mezi průjezdy vozidel.

2.2.3 Rozdělení náhodné veličiny

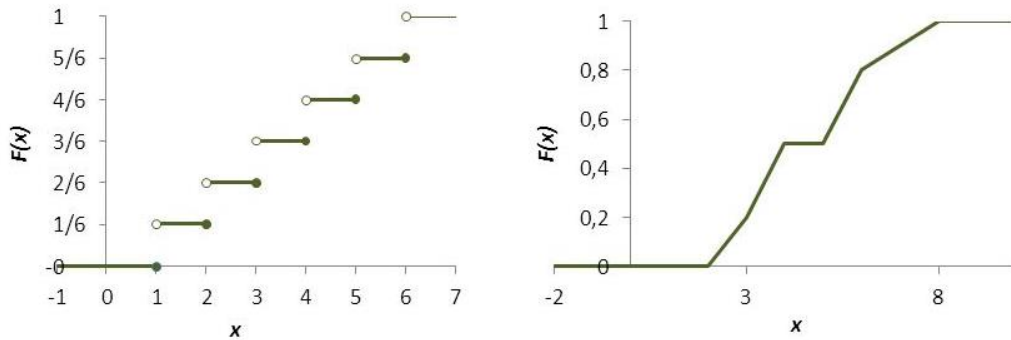
Úplný popis náhodné veličiny se nazývá rozdělení náhodné veličiny. Je to vymezení přípustných hodnot a přiřazení pravděpodobnosti k jednotlivým hodnotám, případně jejich množinám. Toto přiřazení je prováděno pomocí distribuční funkce, pravděpodobnostní funkce nebo hustoty pravděpodobnosti.

Distribuční funkci definuje vztah

$$F_X(x) = P(X \leq x_i)$$

kde X je náhodná proměnná a x_i je libovolné reálné číslo. Matematický zápis distribuční funkce interpretujeme jako pravděpodobnost, že náhodná veličina (proměnná X) nabude hodnoty menší nebo rovno daného x_i . Lze ji použít jak pro diskrétní, tak pro spojitou veličinu. Vždy pro ni platí, že je neklesající v celém oboru, její limita vlevo je 0 a limita vpravo je 1.

Pro ilustraci lze uvést příklad distribuční funkce pro diskrétní a spojitou veličinu na obr. 6. V případě diskrétní veličiny (graf vlevo) můžeme odvodit, že např. pravděpodobnost, že výsledek bude menší nebo roven 4, je 3/6 (tj. 50 %). U spojitě veličiny (graf vpravo) lze určit, že např. pravděpodobnost, že hodnota x bude menší nebo rovna 3, je 0,2 (20 %). [7]

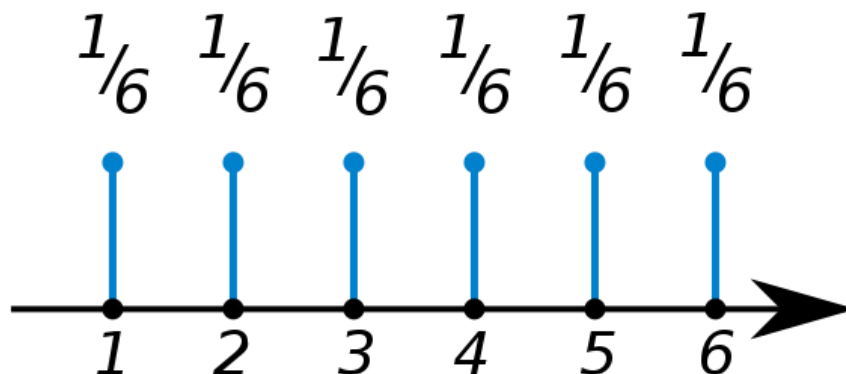


obr. 6 - Distribuční funkce diskrétní a spojité veličiny [7]

Pravděpodobnostní funkce popisuje pouze diskrétní veličinu a je definována vztahem

$$f_X(x_i) = P(X = x_i)$$

kde $f_X(x_i)$ je pravděpodobnostní funkce pro konkrétní hodnotu x_i a $P(X = x_i)$ pravděpodobnost, se kterou daná hodnota x_i nastane. Pro každou realizaci diskrétní veličiny určuje přímo její pravděpodobnost. Všechny členy funkce jsou nezáporné a jejich součet je 1. Konkrétně může nabývat tvaru, který je na obr. 7 a který zobrazuje pravděpodobnostní funkci hrací kostky. Na vodorovné ose jsou všechny možné realizace, na svislé jejich pravděpodobnosti. V případě hrací kostky má všech 6 možných realizací stejnou pravděpodobnost.



obr. 7 - Příklad pravděpodobnostní funkce [8]

Hustota pravděpodobnosti je obdobou pravděpodobnostní funkce pro spojité veličiny. Z definice spojité veličiny, která má nekonečně mnoho realizací, není možné, aby byla definována stejně. Proto je definována za pomoci integrálního počtu

$$F_X(x) = \int_{-\infty}^x f_X(t) dt$$

Hustota pravděpodobnosti je v celém definičním oboru nezáporná a součet plochy pod její křivkou je roven 1.

2.2.4 Sdružená, marginální, podmíněná pravděpodobnost

V předchozích kapitolách byly popisovány náhodné pokusy pouze s jednou náhodnou veličinou a její pravděpodobností. Náhodný pokus ale lze provádět i s více veličinami najednou, kdy při jednom provedení experimentu je jeho realizaci přiřazena hodnota $P(X)$, kde X je vektor náhodných veličin. V tomto případě se jedná o sdruženou pravděpodobnost $P(X)$.

Obdobně jako pravděpodobnost jedné veličiny lze vypočítat i sdruženou pravděpodobnost více veličin. Jedná se o podíl počtu experimentů, kdy nastal daný jev (kombinace požadovaných realizací jednotlivých veličin), vůči celkovému počtu experimentů.

Příkladem může být tab. 1, kde je zaznamenán průběh fiktivního experimentu. Existují v něm 3 zaznamenávané veličiny, každá se 2 možnými realizacemi:

- porušení dopravních předpisů (P)
 - ano (1)
 - ne (0)
- přítomnost alkoholu v krvi řidiče (A)
 - ano (1)
 - ne (0)
- denní doba (D)
 - den (1)
 - noc (0)

tab. 1 - Příklad zaznamenaných veličin [zdroj vlastní]

Číslo	Porušení (P)	Alkohol (A)	Doba (D)
1	1	1	0
2	1	0	1
3	0	0	0
4	1	0	0
5	0	0	1
6	0	1	1
7	0	0	0
8	1	1	1

Úkolem může být zjistit, jaká je pravděpodobnost, že došlo k porušení dopravních předpisů ($P = 1$), bez přítomnosti alkoholu v krvi řidiče ($A = 0$) a v noci ($D = 0$).

Pro tento vzorek dat je sdružená pravděpodobnost $P(P = 1, A = 0, D = 0) = \frac{1}{8}$. Požadovaná kombinace nastala pouze při měření č. 4 z celkového počtu 8 měření.

Dalším pojmem, který bude dále využíván, je marginální pravděpodobnost. Jde o pravděpodobnost výskytu konkrétní náhodné veličiny (tedy např. denní doby) a realizace jejích hodnot bez ohledu na další veličiny.

Pokud bychom chtěli zjistit, jaká je pravděpodobnost, že v naměřeném vzorku dat byl den ($D = 1$),

výsledek je zapsán $P(D = 1) = \frac{4}{8} = \frac{1}{2}$.

Posledním pojmem je podmíněná pravděpodobnost. Ta uvažuje pravděpodobnost jedné nebo více veličin za dané podmínky. Například pravděpodobnost, že u řidiče byla zjištěna přítomnost alkoholu v krvi ($A = 1$) za podmínky, že byla noc ($D = 0$), je zapsána $P(A = 1|D = 0) = \frac{1}{4}$. V tomto případě do jmenovatele vstupují pouze ty vzorky dat, kdy je podmínka $D = 0$ splněna. To znamená, že pravděpodobnost jedné veličiny je ovlivněna pravděpodobností veličiny druhé.

2.3 Bayesovské sítě

Tato kapitola navazuje na poznatky zejména z podkapitol 2.2.3 a 2.2.4 a popisuje další pojmy a jevy, které Bayesovské sítě využívají. Po vysvětlení složitějších teoretických pojmů jsou poznatky shrnuty do ucelené definice Bayesovských sítí. Na definici navazuje i popis algoritmů, pomocí kterých lze Bayesovské sítě tvořit. Princip jejich fungování bude poté ukázán na jednoduchém příkladu. V této kapitole bylo čerpáno ze zdroje [9].

2.3.1 Řetězové pravidlo, Bayesův vzorec

Pro práci s výše uvedenými pravděpodobnostmi (podmíněná, sdružená a marginální) není možné se obejít bez tzv. řetězového pravidla, které vyjadřuje vztah mezi sdruženou, marginální a podmíněnou pravděpodobnostmi. Pro dvě náhodné veličiny A, B má pravidlo tvar

$$P(A, B) = P(A|B) \cdot P(B)$$

Použití řetězového pravidla na sdruženou pravděpodobnostní funkci se nazývá faktorizace sdružené pravděpodobnostní funkce. Nevztahuje se pouze na dvě veličiny, ale lze ho aplikovat na libovolné množství veličin.

Pro příklad uvedený v předchozí kapitole a veličiny porušení dopravních předpisů (P), přítomnost alkoholu v krvi u řidiče (A) a denní doba (D) je vztah

$$P(P, A, D) = P(P|A, D) \cdot P(A|D) \cdot P(D)$$

Pro výpočet sdružené pravděpodobnosti pro kombinaci $P = 1, A = 0, D = 0$ podle tab. 1 vypadá vzorec

$$P(P = 1, A = 0, D = 0) = P(P = 1|A = 0, D = 0) \cdot P(A = 0|D = 0) \cdot P(D = 0) = \frac{1}{3} \cdot \frac{3}{4} \cdot \frac{4}{8} = \frac{1}{8}$$

Využití pravidla nezávisí na pořadí veličin, výsledná pravděpodobnost je stejná

$$P(P = 1, A = 0, D = 0) = P(D = 0|A = 0, P = 1) \cdot P(A = 0|P = 1) \cdot P(P = 1) = \frac{1}{2} \cdot \frac{2}{4} \cdot \frac{4}{8} = \frac{1}{8}$$

Při určování konkrétní pravděpodobnosti se může stát, že tato pravděpodobnost není z nějakého důvodu známa. Jednou z možností jejího vypočtení je potom použití tzv. Bayesova vzorce, které je pro dvě náhodné veličiny definováno jako

$$P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D)}$$

Tímto vztahem poté lze nahradit v rovnici řetězového pravidla neznámé podmíněné pravděpodobnosti.

$$\begin{aligned} P(P, A, D) &= P(P|A, D) \cdot P(A|D) \cdot P(D) = P(P|A, D) \cdot \frac{P(D|A) \cdot P(A)}{P(D)} \cdot P(D) = \\ &= P(P|A, D) \cdot P(D|A) \cdot P(A) \end{aligned}$$

2.3.2 Marginální nezávislost, podmíněná závislost, podmíněná nezávislost

Pro práci z Bayesovskými sítěmi je také nutné znát pojmy, které jsou vysvětleny v této podkapitole. Jedná se o pojmy marginální nezávislost, podmíněná závislost a podmíněná nezávislost. Marginálně nezávislé jsou ty veličiny, o kterých víme nebo předpokládáme, že na sobě nijak nezávisí, tzn. že pravděpodobnost jedné veličiny nijak neovlivňuje pravděpodobnost druhé veličiny. Za předpokladu marginální nezávislosti lze podmíněnou pravděpodobnost zjednodušit

$$P(A|D) = P(A)$$

Místo podmíněné pravděpodobnosti tedy lze uvažovat pouze pravděpodobnost marginální. Stejný vztah platí, i pokud obrátíme proměnné v podmíněné pravděpodobnosti

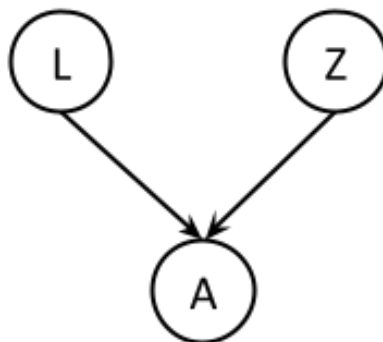
$$P(D|A) = P(D)$$

Pokud např. víme, že proměnné A a D jsou marginálně nezávislé, lze toho využít při výpočtu sdružené pravděpodobnosti.

$$P(A, D) = P(A|D) \cdot P(D) = P(A) \cdot P(D)$$

Dalším pojmem, který je třeba vysvětlit, je podmíněná nezávislost. Uvedeno na příkladu, jehož schéma je zobrazeno na obr. 8, kdy uvažujeme následující situaci. Uzel L reprezentuje náhodnou veličinu loupež, uzel Z náhodnou veličinu zemětřesení, každý uzel má právě 2 stavy (realizace), které mohou nastat, každý z nich s určitou pravděpodobností. Důležité je, že uzly L a Z jsou na sobě nezávislé. Uzel A reprezentuje náhodnou veličinu alarm, který se spustí s určitou (podmíněnou) pravděpodobností podle toho, jestli dochází, nebo nedochází k loupeži nebo k zemětřesení.

Pracujeme s předpokladem, že se dozvíme o spuštění alarmu a zároveň víme, že ve stejné době dochází k zemětřesení. Na základě těchto předpokladů se výrazně sníží pravděpodobnost toho, že dochází i k loupeži. Vztah uzlů L a Z je za podmínky znalosti stavu uzlu A a jednoho z uzlů je vyjádřen jako podmíněná závislost.

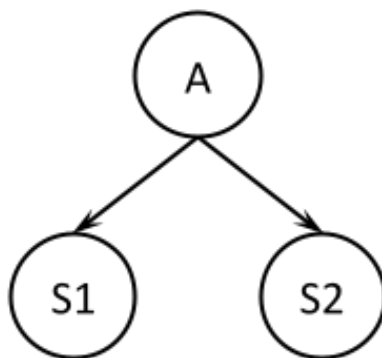


obr. 8 - Schéma příkladu podmíněné závislosti [zdroj vlastní]

Matematicky lze vztah podmíněné závislosti vyjádřit následovně

$$P(A|L) \neq P(A|L, Z)$$

Posledním pojmem, který je nutné vysvětlit, je podmíněná nezávislost. Uvažujeme situaci, která je zobrazena na obr. 9. Uzel A reprezentuje alarm v našem domě se dvěma stavy s danou pravděpodobností (spuštěný/nespuštěný), a uzly S1 a S2 naše dva sousedy, kteří případně spuštěný alarm slyší a na základě toho nám zavolají. Každý ze sousedů má danou pravděpodobnost, se kterou nám v případě spuštěného alarmu zavolá. Zároveň pracujeme s předpokladem, že pokud dojde ke spuštění alarmu, přijde upozornění, na základě kterého budeme mít jistotu, že alarm je spuštěný. Předpokládáme situaci, že dojde ke spuštění alarmu a nám přijde upozornění, máme tedy jistotu, že alarm opravdu spuštěný je. Zároveň s tím se zvýší pravděpodobnost hovoru u obou sousedů. Jakmile ale dojde k tomu, že zavolá soused S1, pravděpodobnost, že zavolá soused S2 se nezmění. Právě skutečnost, že pravděpodobnost hovoru od druhého souseda za podmínky, že víme o spuštění alarmu, se nezmění, je důkazem podmíněné nezávislosti.



obr. 9 - Schéma pro příklad podmíněné nezávislosti [zdroj vlastní]

Matematicky vypadá zápis podmíněné nezávislosti při využití proměnných z příkladu následovně.

$$P(A|S1, S2) = P(A|S2)$$

Tohoto vztahu lze využít při zjednodušení výpočtu sdružené pravděpodobnosti. Pokud budeme předpokládat, že uzly S1 a S2 jsou podmíněně nezávislé, lze původní vztah zjednodušit.

$$P(A, S1, S2) = P(A|S1, S2) \cdot P(S1|S2) \cdot P(S2) = P(A|S2) \cdot P(S1|S2) \cdot P(S2)$$

2.3.3 Definice Bayesovské sítě

V návaznosti na již popsané pojmy lze úplně definovat Bayesovské sítě. Bayesovská síť představuje příčinný pravděpodobnostní vztah mezi souborem náhodných veličin, jejich podmíněné závislosti a je reprezentací sdruženého rozdělení pravděpodobnosti. Tvoří ji struktura definovaná orientovaným acyklickým grafem a množina lokálních pravděpodobnostních rozdělení. V orientovaném acyklickém grafu jednotlivé uzly reprezentují proměnné, hrany existenci jejich vzájemné závislosti a orientace hran kauzalitu mezi nimi. Každému uzlu je přiřazeno lokální pravděpodobnostní rozdělení, které je podle struktury buď marginální, nebo podmíněné. Platí, že jak struktura, tak lokální pravděpodobnostní rozdělení mohou být určeny výpočtním algoritmem i expertně. [10]

Existuje obecný vzorec používaný pro výpočet sdružené pravděpodobnostní funkce všech proměnných. Tento vzorec vychází z výpočtu sdružené pravděpodobnosti a jejího zjednodušení na základě marginální nezávislosti.

$$P(A, B, C) = P(A|B, C) \cdot P(B|C) \cdot P(C) = P(A|B, C) \cdot P(B) \cdot P(C)$$

Po zobecnění vztahu popsaného výše pro všechny proměnných lze obecný vztah pro vektor proměnných \mathbf{X} , kde $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$, zapsat

$$P(\mathbf{X}) = \prod_{i=1}^N P(x_i | Pa(x_i))$$

Jedná se o součin jednotlivých proměnných, přičemž pro každou proměnnou je vypočtena podmíněná pravděpodobnost proměnné x_i za podmínky rodičů dané proměnné $Pa(x_i)$. Jinými slovy je provedena faktorizace sdružené pravděpodobnosti na lokální pravděpodobnosti, které obsahují pouze uzel a jeho rodiče. Tento vztah je založen na předpokladu, že každý uzel v Bayesovské síti je podmíněně nezávislý na všech uzlech, které nejsou jeho následníky, vzhledem k jeho rodičům. V praxi to znamená, že pro lokální pravděpodobnostní rozdělení každé proměnné (každého uzlu) stačí brát v potaz pouze její rodiče, což výrazně snižuje výpočetní náročnost. Zároveň z tohoto vyplývá, že pokud daná proměnná (daný uzel) nemá rodiče, lze u nich podmíněnou pravděpodobnost nahradit marginální.

2.3.4 Algoritmus pro tvorbu sítě

Obecně existují 2 hlavní postupy, jak vytvořit Bayesovskou síť, tedy její strukturu (graf) a lokální distribuce pro všechny uzly. Jedním je vytvoření expertem (experty), kteří síť tvoří na základě své odbornosti a zkušenosti. Tento postup byl v práci využit na příkladech v podkapitolách 2.3.2 a 2.3.5. Druhým postupem je vytvoření sítě pomocí algoritmu na základě dat. Ten byl využit v praktické části této práce. Předmětem dalších vědeckých prací je i kombinace těchto 2 postupů za pomoci slučovacího algoritmu. [11]

Existují 2 základní druhy algoritmů, které jsou nejčastěji používány pro tvorbu struktury sítě:

- score-based algoritmy (založené na skóre)
- constraint-based algoritmy (založené na omezení)

Score-based algoritmy pracují s ohodnocením tvořené struktury sítě. Definují skóre, které hodnotí, jak dobře závislosti nebo nezávislosti ve struktuře odpovídají datům, a hledá strukturu, která maximalizuje skóre. Problémem je množství různých variant, které algoritmus porovnává. Existují ale metody, které dokáží udržet počet na zvládnutelné úrovni.

Constraint-based algoritmy tvoří síť podle vztahů mezi vstupními proměnnými. Provádí testy podmíněné nezávislosti na datech a hledá síť, která je v souladu se zjištěnými závislostmi a nezávislostmi. Pracují na základě 2 předpokladů: že se podařilo změřit všechny hodnoty všech proměnných a že každá proměnná je podmíněně nezávislá na všech proměnných, které nejsou jejími

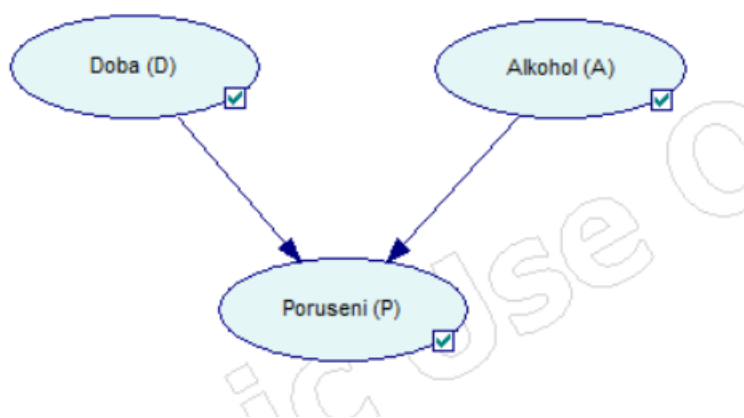
následníky, vzhledem k jejím rodičům. Nevýhodou je nedostatečná robustnost, kdy malá změna na začátku může mít velký vliv na konečný výsledek.

Výhodou score-based algoritmů oproti constraint-based algoritmům je, že jsou méně citlivé na chyby v jednotlivých testech. Lze dosáhnout kompromisu mezi mírou závislosti proměnných v datech a náklady na přidání hrany. V této práci je pro tvorbu struktury sítě použit hill-climbing algoritmus, který patří mezi score-based algoritmy. Hill-climbing algoritmus patří mezi rychlé a široce využívané algoritmy, který zároveň poskytuje dobré výsledky z hlediska složitosti a přesnosti sítě. [12]

2.3.5 Příklad Bayesovské sítě a práce s ní

Na závěr kapitoly bude na jednoduchém příkladu představena Bayesovská síť a práce s ní. K sestavení sítě a práci s ní byl využit software GeNIe. V tomto softwaru je možné síť vytvářet, graficky zobrazovat i modelovat. Jedná se o software volně dostupný pro výukové a vědecké účely, který splňuje všechny požadavky spojené s touto prací – grafické zobrazení topologie sítě nebo přehledné zobrazení lokálních pravděpodobnostních funkcí pro každý uzel. Informace k softwaru, dokumentace a ukázky z používání jsou dostupné na odkazu: <https://www.bayesfusion.com/genie/>.

Jako ukázková síť byla vybrána síť vycházející z veličin použitých ve fiktivním experimentu v podkapitole 2.2.4. Obsahuje 3 uzly – porušení dopravních předpisů (P), přítomnost alkoholu v krvi řidiče (A), doba (D), které tvoří množinu proměnných. Odpovídá tomu sdružená pravděpodobnostní funkce $P(P, A, D)$. Každý z uzlů má 2 stavy, ke kterým byly přiřazeny příslušné pravděpodobnosti. Tyto pravděpodobnosti byly přiřazeny autorem a nebyly získány žádným měřením, jsou pouze reprezentativní pro tento příklad. Síť je zobrazena na obr. 10.



obr. 10 - Příklad Bayesovské sítě vytvořené v softwaru GeNIe [zdroj vlastní]

Z obrázku je patrné, že uzly D, A nemají žádné rodiče, bude jim tedy přidělena marginální pravděpodobnost $P(D)$, resp. $P(A)$. V souladu s poznatky z předchozích kapitol lze o uzlech D, A říct, že se jedná o marginálně nezávislé proměnné. Za podmínky znalosti stavu uzlu P se uzly stávají podmíněně závislými. Hodnoty marginálních pravděpodobností pro oba zmíněné uzly jsou v tab. 2 a tab. 3.

tab. 2 - Marginální pravděpodobnost $P(D)$ [zdroj vlastní]

Doba	D	$P(D)$
Noc	0	0,33
Den	1	0,67

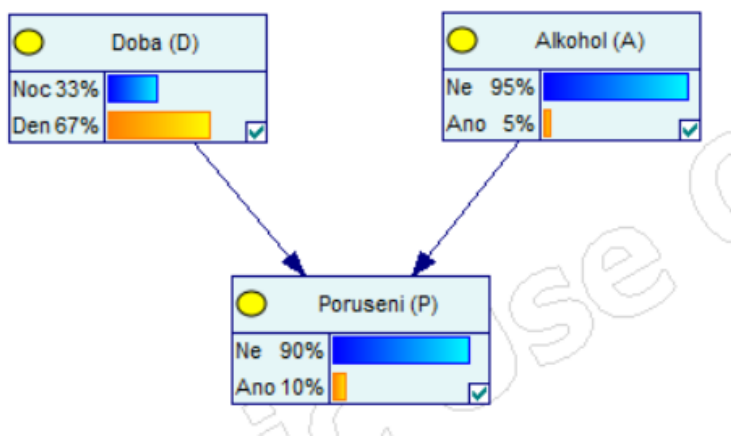
tab. 3 - Marginální pravděpodobnost $P(A)$ [zdroj vlastní]

Alkohol	A	$P(A)$
Ne	0	0,95
Ano	1	0,05

Uzel P je dítětem uzlů D, A a musí mu tedy být přiděleny podmíněná pravděpodobnost $P(P|D, A)$. Ta je zobrazena v tab. 4. Na obr. 11 je síť zobrazena graficky s pravděpodobnostmi.

tab. 4 - Podmíněná pravděpodobnost $P(P|D, A)$ [zdroj vlastní]

Doba			D = 0		D = 1	
Alkohol			A = 0	A = 1	A = 0	A = 1
Porušení dopravních předpisů	Ne	0	0,85	0,4	0,95	0,7
	Ano	1	0,15	0,6	0,05	0,3



obr. 11 - Příklad sítě s pravděpodobnostmi [zdroj vlastní]

Je patrné, že hodnoty pravděpodobnosti v uzlu P neodpovídají žádným hodnotám z tab. 4. Je to proto, že v tabulce je zadána podmíněná pravděpodobnost $P(P|D,A)$, ale v grafickém zobrazení sítě v softwaru GeNIe už se jedná o vypočtenou marginální pravděpodobnost $P(P)$, která platí pouze pro uzel P.

Pro výpočet marginální pravděpodobnosti je využita tzv. marginalizace. Jde o přepočítání, který využívá vztahu

$$P(P) = \sum_{A,D} P(P|D,A) \cdot P(D,A)$$

První činitel součinu $P(P|D,A)$ je podmíněná pravděpodobnost již známá z tab. 4. Druhý činitel je sdružená pravděpodobnost uzlů D, A. Z předchozích kapitol víme, že uzly D, A jsou marginálně nezávislé, lze tedy sdruženou pravděpodobnost upravit na součin pravděpodobností marginálních

$$P(D,A) = P(D) \cdot P(A)$$

Na základě tohoto vztahu lze vypočítat sdruženou pravděpodobnost pro všechny kombinace hodnot uzlů D, A. Tyto kombinace jsou zobrazeny v tabulce tab. 5.

tab. 5 - Sdružená pravděpodobnost $P(D,A)$ [zdroj vlastní]

Doba	0	1
Alkohol		
0	$0,33 \cdot 0,95 = 0,3135$	$0,67 \cdot 0,95 = 0,6365$
1	$0,33 \cdot 0,05 = 0,0165$	$0,67 \cdot 0,05 = 0,0335$

Výpočet konkrétní hodnoty v uzlu P, např. $P(P = 1)$, vypadá následovně

$$\begin{aligned} P(P = 1) &= P(P = 1|D = 0, A = 0) \cdot P(D = 0, A = 0) + \\ &P(P = 1|D = 1, A = 0) \cdot P(D = 1, A = 0) + \\ &P(P = 1|D = 0, A = 1) \cdot P(D = 0, A = 1) + \\ &P(P = 1|D = 1, A = 1) \cdot P(D = 1, A = 1) \end{aligned}$$

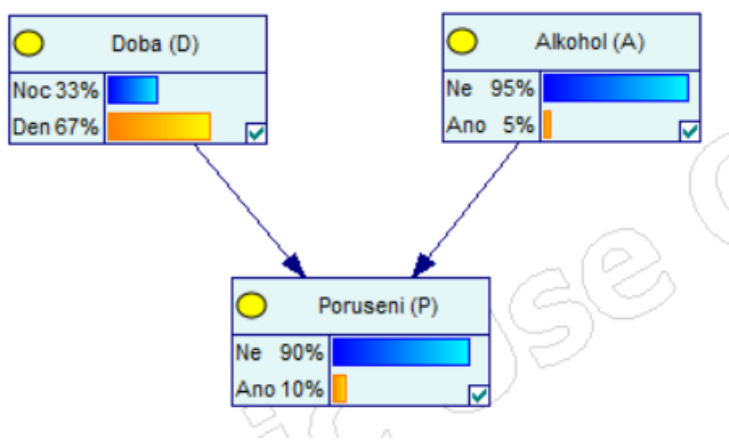
Poté stačí jednotlivé hodnoty do vztahu dosadit a vypočítat. Obdobným způsobem probíhá výpočet pro $P(P = 0)$. Výsledné marginální pravděpodobnosti $P(P)$ jsou zobrazeny v tab. 6. Ověřením správnosti postupu a výpočtu je to, že výsledek se shoduje s výpočtem softwaru, který je na obr. 11.

tab. 6 - Vypočtená marginální pravděpodobnost $P(P)$ [zdroj vlastní]

Porušení	P	$P(P)$
Ne	0	$0,85 \cdot 0,3135 + 0,95 \cdot 0,6365 + 0,4 \cdot 0,0165 + 0,7 \cdot 0,0335 = 0,9012 \doteq \mathbf{90 \%}$
Ano	1	$0,15 \cdot 0,3135 + 0,05 \cdot 0,6365 + 0,6 \cdot 0,0165 + 0,3 \cdot 0,0335 = 0,0988 \doteq \mathbf{10 \%}$

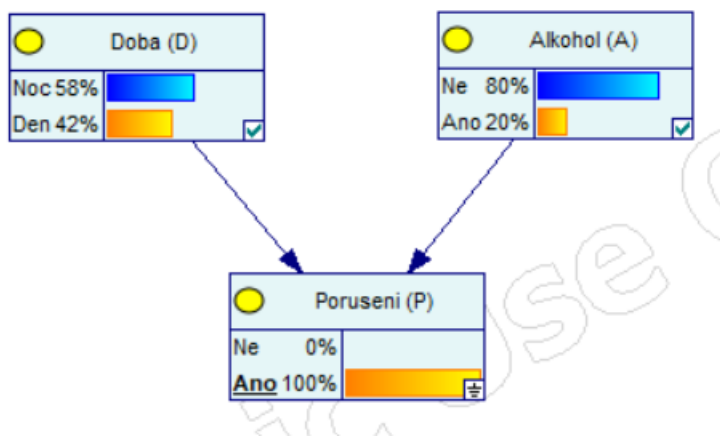
Důležitou součástí práce s Bayesovskými sítěmi je tzv. inference. Inference v Bayesovských sítích spočívá ve výpočtu podmíněné pravděpodobnosti některých proměnných za předpokladu, že pro

jednu nebo více proměnných je zavedena evidence. Inference může být provedena pro určitý stav nebo hodnotu proměnné za předpokladu důkazů o stavu ostatních proměnných. Zavedením evidence je myšlena situace, kdy pravděpodobnost realizace určité hodnoty určité náhodné veličiny je 1 (100 %), tedy je jisté, že vždy nastane pouze tato určitá hodnota a ostatní nastat nemůžou. Inference poté znamená přepočtení pravděpodobnostního rozdělení pro ostatní uzly s využitím zavedené evidence. [13] Při využití výše uvedeného příkladu lze ilustrovat zavedení evidence do uzlu sítě a následnou inferenci. Na obr. 12 je zobrazena síť s 3 uzly a pravděpodobnostmi zadanými expertem, jak již byla popsána v předchozích kapitolách.



obr. 12 - Síť před zavedením evidence [zdroj vlastní]

Na obr. 13 je zobrazena síť po zavedení evidence *ano* v uzlu P (porušení dopravních předpisů). Jedná se o stav, kdy víme s jistotou, že došlo k porušení dopravních předpisů, je tedy nastavena pravděpodobnost 100 % pro stav *ano* a pravděpodobnost 0 % pro stav *ne*. Vzhledem k existujícím vztahům se zbývajícími 2 uzly se zavedení evidence projeví i na pravděpodobnostech jejich stavů.



obr. 13 - Síť po zavedení evidence [zdroj vlastní]

Z obrázku vyplývá, že pokud dojde k porušení dopravních předpisů ($P = ano$), zvýšila se pravděpodobnost pro stav $D = noc$ z původních 33 % na 58 % a pro stav $A = ano$ z 5 % na 20 %.

2.4 Shrnutí teoretické kapitoly

V předchozích kapitolách byly představeny základní pojmy z teorie grafů a teorie pravděpodobnosti, které jsou nezbytné pro pochopení fungování Bayesovských sítí a práci s nimi. Z teorie grafů se jedná o pojmy, které vedly k definici typu grafu používaného v Bayesovských sítích – orientovaný acyklický graf. Z teorie pravděpodobnosti to byly marginální, podmíněná a sdružená pravděpodobnost, na kterou navázaly poznatky o Bayesově vzorci, podmíněné závislosti a nezávislosti a marginální nezávislosti. Kde to bylo možné a vhodné, pojmy a vzorce byly doplněny konkrétními příklady a výpočty, aby bylo možné princip lépe představit.

V závěru kapitoly byly Bayesovské sítě definovány za pomoci již popsaných pojmů, byly představeny algoritmy využívané při tvorbě struktury sítí a na jednoduchém příkladu byl představen i princip výpočtu marginálního pravděpodobnostního rozdělení v Bayesovské síti. V rámci příkladu byly definovány také pojmy inference a zavedení evidence v síti.

3 Zpracování nehodových dat

V této kapitole je popsána aplikace teoretických poznatků z předchozích kapitol na reálná nehodová data. V úvodu této kapitoly jsou stručně popsány výhody Bayesovských sítí pro analýzu nehodových dat, dále je charakterizována oblast, ze které pochází zpracovávaná data. Tato charakteristika se týká zejména dopravního provozu a vývoje statistiky nehodovosti. V dalších podkapitolách je popsán datový soubor, postup při jeho zpracování. Vysvětleny jsou úpravy jednotlivých zaznamenávaných parametrů, aby mohla být vytvořena Bayesovská síť. V rámci tvorby Bayesovské sítě je popsána vytvořená síť z hlediska uzlů a vazeb a výstupy, které se týkají zavedení evidence do uzlu vážnost.

3.1 Bayesovské sítě a datová analýza

Existuje nespočet metod a nástrojů, pomocí kterých lze datovou analýzu provádět. Pokud se jedná o úlohu zjištění uspořádání vztahů mezi velkým počtem proměnných, jedním z nevhodnějších nástrojů jsou bezesporu Bayesovské sítě.

Hlavním účelem Bayesovských sítí je najít možné příčiny vybraného měřeného výsledku. Běžným přístupem tohoto druhu analýzy dat je vytvoření Bayesovské sítě pomocí algoritmu založeného na datech a následný výpočet Bayesovské inference za předpokladu předchozích znalostí. Výhody Bayesovských sítí jsou především v tom, že se jedná o silný modelovací nástroj pro datovou analýzu, který se opírá i graficky přehlednější znázornění situace.

Úloha datové analýzy týkající se dat o dopravních nehodách k využití Bayesovských sítí vybízí. U každé dopravní nehody je možné rozlišit velké množství faktorů, které ke vzniku nehody mohly přispět. Právě jejich množství a možné vztahy mezi nimi jsou výzvou k modelování pomocí Bayesovských sítí.

3.2 Datový soubor

Kapitola nejprve krátce popisuje software RStudio, který byl použit ke zpracování a přípravě datového souboru, vytvoření Bayesovské sítě a následnou práci s ní. Dále jsou představeny oblasti, ze kterých nehodová data pochází, z hlediska obecné i dopravní charakteristiky. Jsou uvedeny i počty dopravních nehod a další ukazatele týkající se dopravních nehod za zpracovávané období.

V další části kapitoly je popsána příprava datového souboru pro tvorbu Bayesovské sítě. Kromě popisu vyřazených nebo sloučených parametrů spolu s odůvodněním jde i o úpravu jednotlivých stavů v rámci jednoho parametru. Zároveň jsou pro lepší představu o datech uvedeny i četnosti jednotlivých stavů.

3.2.1 Použitý software

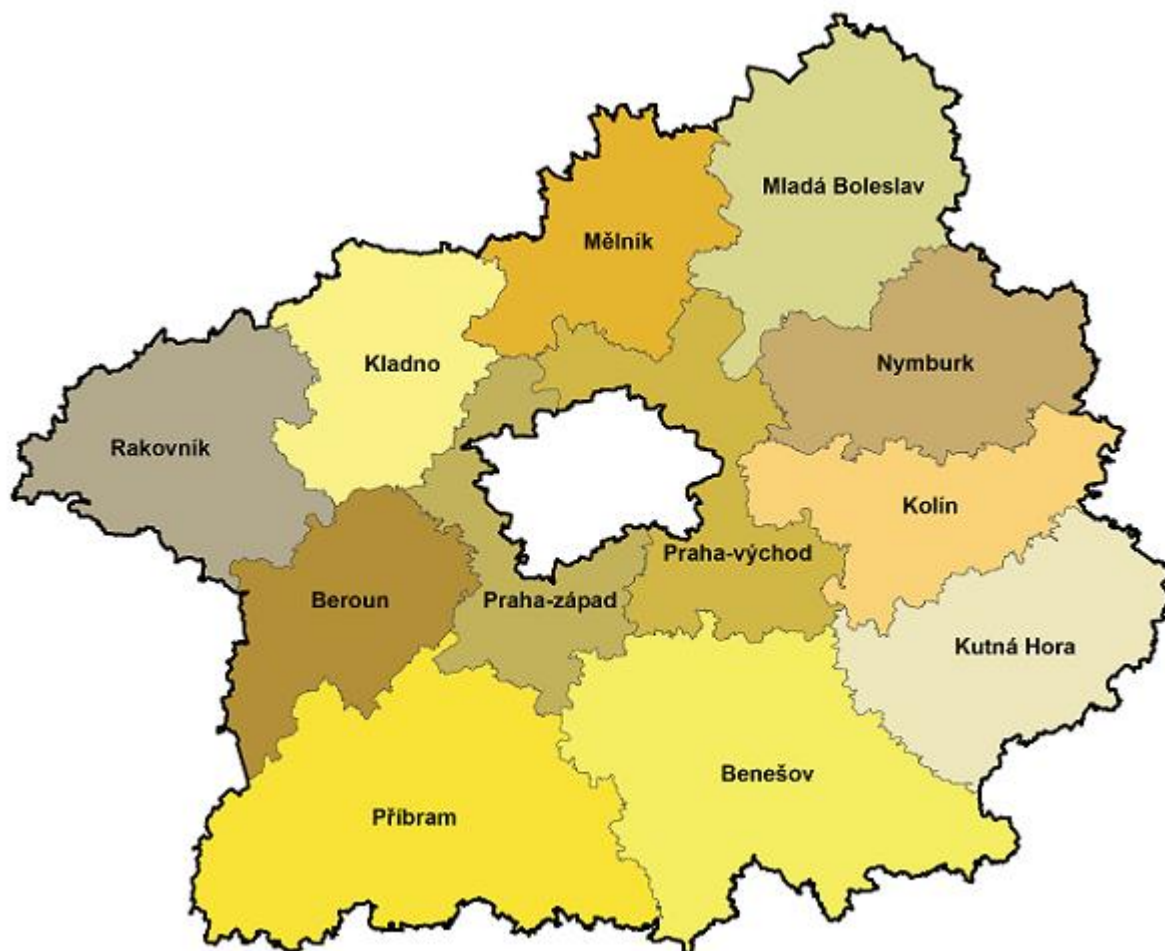
Data byla zpracovávána v RStudios, což je volně dostupný software umožňující práci s programovacím jazykem R. R je jazyk a prostředí pro statistické výpočty a grafiku, který poskytuje širokou škálu statistických (lineární a nelineární modelování, klasické statistické testy, analýza časových řad, klasifikace, shlukování, ...) a grafických technik. Pro rozšířené možnosti zpracování dat je možné instalovat balíčky (packages), které obsahují další funkce. Mezi balíčky, které byly využity v rámci této práce, patří mj. bnlearn, Rgraphviz nebo gRain, které umožňují tvorbu Bayesovských sítí a následnou práci s nimi. Celý kód je součástí příloh ve formátu .R jako příloha 1. [14]

3.2.2 Charakteristika zpracovávané oblasti

Význam této podkapitoly spočívá v poskytnutí uceleného přehledu oblasti. Uvedeny jsou základní údaje o daných územích, dále jejich charakteristika z hlediska dopravního provozu a vybrané statistické údaje.

Vybraná nehodová data představují záznamy o nehodách z let 2011-2021 na území hl. m. Prahy a okresů Praha-východ a Praha-západ. Jedná se pouze o nehody, které zaznamenala Policie ČR. Zdrojem dat je Policejní prezidium ČR.

Na obr. 14 je zobrazena poloha dotčených okresů Praha-východ a Praha-západ v rámci Středočeského kraje. Hl. m. Praha je na mapě vyznačeno bez popisu bílou barvou.



obr. 14 - Mapa Středočeského kraje [15]

Praha je hlavním městem České republiky. Jedná se o území o velikosti 496 km², které obývá 1,275 milionu obyvatel. Praha je centrálním bodem dopravních vazeb jak pro sousední Středočeský kraj, tak pro území ČR (dojíždka za prací a do škol, za nákupy, kulturou nebo do zdravotnických zařízení). V automobilové dopravě České republiky zaujímá hlavní město Praha zvláštní postavení, které se projevuje v nadprůměrně vysokých dopravních výkonech i intenzitách ve srovnání s jinými českými městy nebo s dálnicemi a silnicemi v extravilánu. [16][18]

V tab. 7 jsou uvedeny vybrané dopravní údaje o hl. m. Praze. Údaje pochází z ročenky dopravy Technické správy komunikací hl. m. Prahy (TSK) za rok 2021, jsou tedy platné k 31.12.2021. [17]

tab. 7 - Vybrané údaje o hl. m. Praze k 31.12.2021 [17]

Celková délka komunikační sítě	4067 km
Počet motorových vozidel	1 217 344
z toho počet osobních automobilů	970 755
Stupeň motorizace (vozidel na 1 000 obyvatel)	955
Počet obyvatel na 1 motorové vozidlo	1,0
Stupeň automobilizace (osobních automobilů na 1 000 obyvatel)	762
Počet obyvatel na 1 osobní automobil	1,3
Dopravní výkony automobilové dopravy na celé komunikační síti	
za průměrný pracovní den	23,0 mil. vozokm
za rok	7,2 mld. vozokm

Okres Praha-východ svou rozlohou 755 km² zaujímá okres 7. místo ve Středočeském kraji a 6,9 % z jeho rozlohy. Charakteristická podoba okresu, tvořící svým excentrickým tvarem část kruhu okolo Prahy, má výrazně protáhlý tvar ve směru sever-jih. Jeho netradiční vzhled, nejzřetelnější v grafické podobě na mapě, dokládají údaje o vzdálenostech. Zatímco z jižního cípu okresu na jeho severní protějšek je přibližně 80 km, na nejúžším místě okresu v katastru obce Úvaly činí jeho šířka pouhé 3 km. Okresem prochází řada významných silničních tahů propojujících Prahu s ostatními částmi republiky - dálnice D1, D8, D10 a D11 a silnice I. třídy I/2, I/3, I/9 a I/12. Stejně tak procházejí okresem i hlavní dálkové železniční tahy vedoucí přes Prahu – tranzitní železniční koridory (I., III. a IV.) propojující republiku se zahraničím. [19]

Okres Praha-západ svou rozlohou 580 km² je okres Praha-západ nejmenším okresem ve Středočeském kraji, zaujímá pouze 5,3 % z jeho rozlohy. Území okresu tvoří půlměsíc obepínající ze západu hlavní město Prahu. Okresem prochází významné silniční tahy vedoucí do Prahy ze západu a z jihu. Dálnice D1, D4, D5, D6 a D7 a silnice I. třídy I/4. Územím prochází III. tranzitní železniční koridor a železniční tratě propojující oblasti Středočeského kraje (Kladensko, Rakovnicko, Berounsko) s Prahou. [20]

Oba okresy mají okres v republice naprosto výjimečné postavení, které dokládá většina sociálněekonomických charakteristik. Tyto dva okresy, uzavírající ve svém středu hlavní město Prahu – mezinárodní metropoli – tvoří spolu pražskou středočeskou aglomeraci a slouží Praze jako její zázemí. Jsou pro Prahu zdrojem pracovních sil, doplňují pražský průmysl, stavebnictví a služby, zásobují Prahu potravinami, poskytují Praze svůj rekreační potenciál. Naopak na území obou okresů probíhá od poloviny 90. let minulého století masivnější výstavba především rodinných domků. Probíhá zde proces tzv. suburbanizace. Pražské ekonomicky silné obyvatelstvo se stěhuje do zázemí metropole s cílem zlepšení kvality bydlení. V důsledku tohoto suburbanizačního procesu dochází v obou okresech k významnému nárůstu počtu obyvatel migrací.

3.2.3 Vývoj nehodovosti

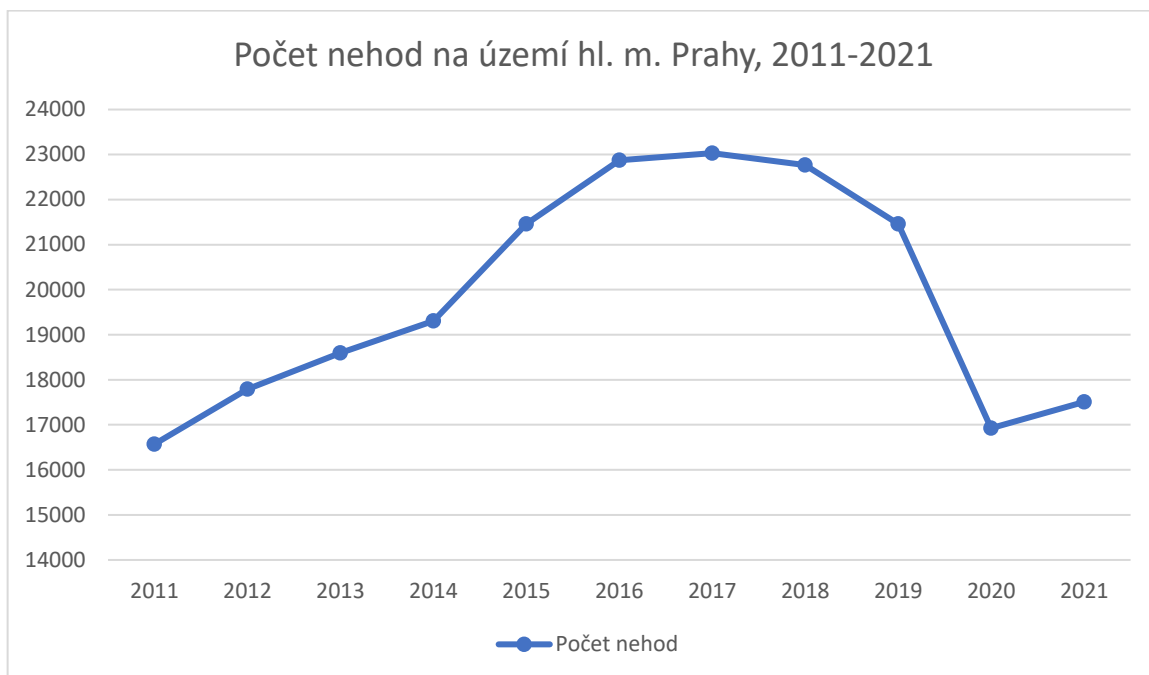
Trend nehodovosti na území hl. m. Prahy se v posledních letech vyvíjel příznivě, nezanedbatelnou roli v tom ovšem hrála pandemie onemocnění covid-19 a omezení s tím spojená zejména v letech 2020 a 2021. Měnila se také legislativa, která souvisí s povinným nahlašování dopravních nehod policii. Celkově lze sledovat příznivý trend zejména v poklesu počtu smrtelně, těžce i lehce zraněných osob, navzdory zvyšujícímu se automobilovému provozu.

V tab. 8 jsou uvedeny počty nehod, počty smrtelně, těžce a lehce zraněných a relativní nehodovost za roky 2011-2021, které jsou zároveň obdobím zpracovávaným v této práci. Relativní nehodovost je vypočítána jako počet nehod na jeden milion ujetých vozokilometrů.

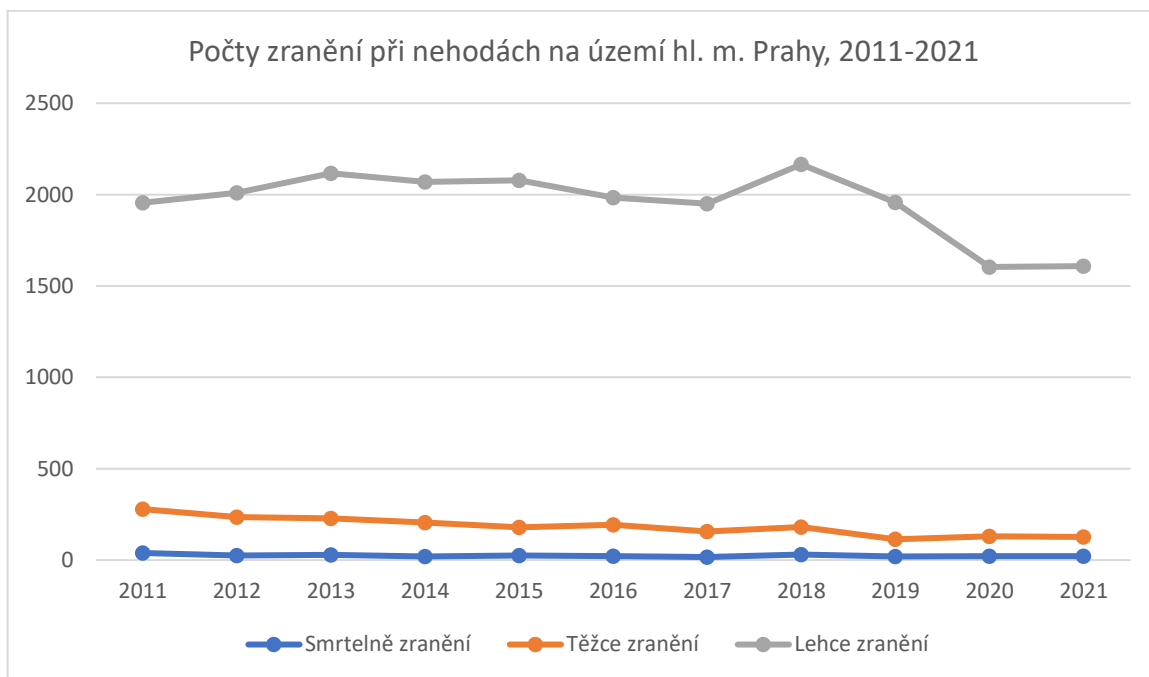
tab. 8 - Počet nehod, zranění a relativní nehodovost v Praze, 2011-2021 [17]

Rok	Počet nehod	Smrtelně zranění	Těžce zranění	Lehce zranění	Rel. nehodovost
2011	16572	39	279	1955	2,3
2012	17795	26	236	2009	2,5
2013	18593	29	228	2116	2,6
2014	19306	20	206	2070	2,7
2015	21462	25	179	2078	3,1
2016	22876	21	194	1983	3,3
2017	23032	17	156	1951	3,2
2018	22767	31	182	2165	3,1
2019	21458	20	114	1958	3,0
2020	16925	22	131	1604	2,5
2021	17510	22	126	1609	2,4

Graficky je vývoj počtu nehod a počtu zranění podle typu za zpracovávané období zobrazují graf 1 a graf 2.



graf 1 - Počet nehod na území hl. m. Prahy, 2011-2021 [17]



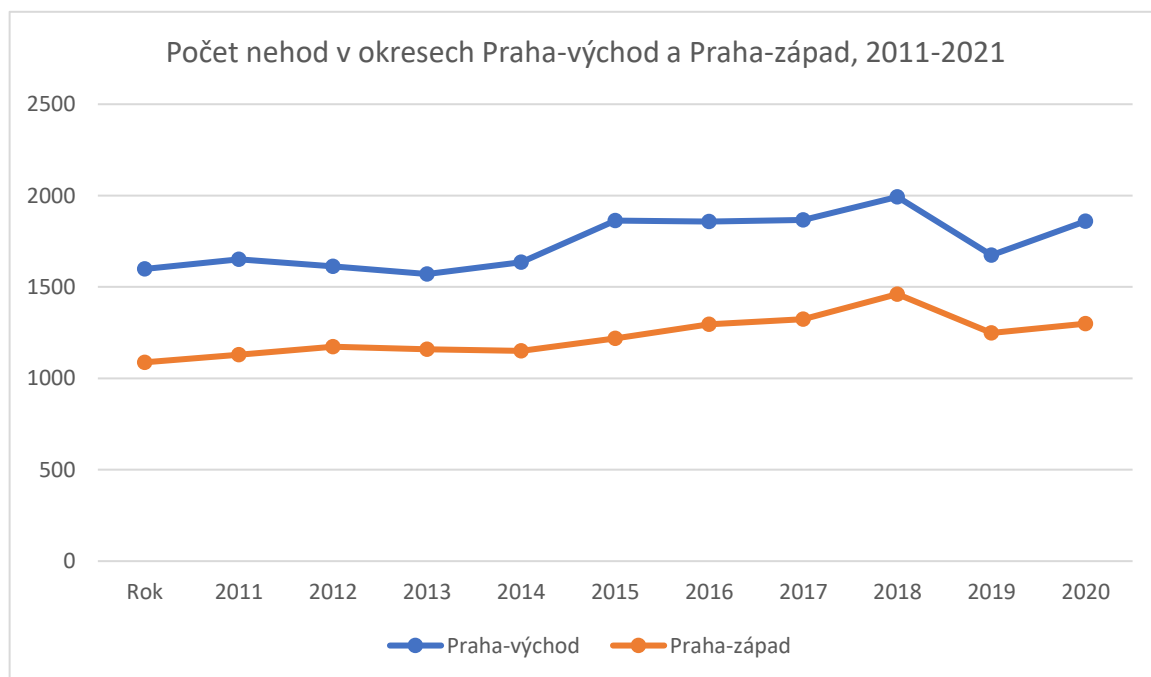
graf 2 - Počty zranění při nehodách na území hl. m. Prahy, 2011-2021 [17]

Z webu Centra dopravního výzkumu pochází počty nehod v okresech Praha-východ a Praha-západ za roky 2011-2021, které jsou shrnuty v tab. 9 a zobrazuje je graf 3. Z tabulky i grafu je patrné, že počet nehod je výrazně nižší než na území hl. m. Prahy za stejné období. Při porovnání mezi oběma okresy je počet nehod v okrese Praha-východ zhruba o 500-600 nehod ročně vyšší. [2]

Z uvedeného zdroje ani z dalších nebylo možné zjistit další údaje jako počty zraněných nebo relativní nehodovost, proto uvedeny nejsou.

tab. 9 - Počty nehod v okresech Praha-východ a Praha-západ, 2011-2021 [2]

Rok	Počet nehod	
	Praha-východ	Praha-západ
2011	1598	1088
2012	1651	1130
2013	1612	1174
2014	1571	1160
2015	1636	1151
2016	1863	1219
2017	1858	1295
2018	1866	1323
2019	1993	1461
2020	1674	1248
2021	1859	1300



graf 3 - Počet nehod v okresech Praha-východ a Praha-západ, 2011-2021 [2]

3.2.4 Popis datového souboru

Data byla od zdroje, kterým je Policejní prezidium ČR, převzata ve formátu .XLSX ve dvou souborech. První soubor obsahuje nehody na území hl. m. Prahy, které jsou rozděleny po letech v jednotlivých

listech (celkem tedy 11 listů). Druhý soubor obsahuje nehody na území okresů Praha-východ a Praha-západ, v jednom listu za celé období.

Každý řádek reprezentuje konkrétní nehodu, která byla zaznamenána podle formuláře pro evidenci nehod v silničním provozu. V průběhu sledovaného období byly do formuláře doplněny další sledované parametry nehody, které ji přesněji popisují, jednotlivé roky se tedy liší počtem těchto zaznamenaných parametrů. První změna proběhla v roce 2016, kdy bylo přidáno 26 nových parametrů, druhá změna potom v roce 2018, kdy bylo doplněno ještě dalších 7 parametrů. V tab. 10 je shrnut celkový počet zaznamenávaných parametrů pro každou nehodu pro výše popsaná období.

tab. 10 - Počet zaznamenávaných položek pro jednotlivá období [zdroj vlastní]

Období	2011-2015	2016-2017	2018-2021
Počet parametrů	56	82	89

Vzhledem k rozdílnému počtu parametrů v jednotlivých letech autor rozhodl, že parametry přidané v pozdějších letech budou vynechány, aby všechny záznamy o nehodách využitě pro tvorbu sítě měly stejnou strukturu. Další parametry byly sloučeny nebo byly redukovány jejich stavy. Vyřazené, upravené i redukované parametry budou popsány v další kapitole.

Komplikací byl také formát dat, který byl po převedení datových souborů do RStudia načten jako písmena, ačkoliv se jednalo o čísla. Tato skutečnost však byla lehce opravitelná změnou formátu daného sloupce.

Důležitou okolností, kterou je nutné vzít v potaz, je fakt, že záznam parametrů nehody závisí na konkrétním policistovi, který vyplňuje příslušný formulář. V závislosti na zkušenostech a dalších okolnostech by tedy některé parametry, jejichž hodnota není zcela zřejmá, mohly být zaznamenány jinou osobou jinak. To by následně ovlivnilo i síť z hlediska struktury a lokálních distribucí vypočtených algoritmem, jelikož by se jednalo i jiná vstupní data.

3.2.5 Vyřazené a sloučené parametry

Jak bylo zmíněno v předchozí kapitole, parametry, které nebyly dostupné pro nehody v celém zpracovávaném období, byly z tvorby sítě vyřazené.

Jedná se o následující parametry:

- p5b – kód obce
- p29 – kategorie chodce
- p30 – stav chodce
- p31 – chování chodce
- p32 – situace v místě nehody s chodcem
- p33c – pohlaví chodce

- p33d – rok narození chodce
- p33e – státní příslušnost chodce
- p33f – první pomoc chodci
- p33g – následky chodce
- p33h – následné úmrtí chodce (dny)
- p40 – číslo x uzel 1 – týká se křižující komunikace
- p41 – km x uzel 2 – týká se křižující komunikace
- p45a – výrobní značka motor. vozidla
- p47 – rok výroby vozidla
- p48a – charakteristika vlastníka vozidla
- p51 – způsob vyproštění osob z vozidla
- p59e – státní příslušnost osoby ve vozidle
- p59h – následné úmrtí osoby ve vozidle (dny)
- f, g, h, i, k, l, m, n, o, p, q, r, s, t – další popis sledované komunikace

Další parametry byly vyřazeny, ačkoliv byly zaznamenávány v celém zpracovávaném období. Jednalo se o parametry, které nebyly z různých důvodů vhodné pro zpracovávaný záměr práce – tvorbu Bayesovské sítě. Mezi tyto důvody patří to, že se jedná o parametry, které na vážnost nehody nemají vliv (ID vozidla), jsou využitelné spíše pro jinak zaměřenou práci (Křovákovy souřadnice), byly obsaženy v podobném parametru (škoda na vozidle – místo ní byl využit parametr celková škoda) nebo jejich záznam nebyl kompletní – často se opakovala neznámá hodnota NULL (parametry p59a – p59g týkající se následků ve vozidle). O tom, které parametry budou vyřazeny, rozhodl autor na základě svého uvážení a následných konzultací s vedoucími práce.

Jedná se o následující parametry:

- p34 – počet zúčastněných vozidel
- p37 – číslo komunikace
- p38 – km
- p39 – druh křižující komunikace
- id – id_vozidla
- p45b – údaje o vozidle
- p48b – doplňující údaje o vozidle
- p50a – vozidlo po nehodě
- p52 – směr jízdy nebo postavení vozidla
- p53 – škoda na vozidle
- p59a – osoba

- p59b – bližší označení osoby
- p59c – pohlaví
- p59d – rok narození
- p59f – první pomoc
- p59g – následky
- X, Y – Křovákovy souřadnice

Další kategorií parametrů byly parametry, které byly na základě pravidla nebo vzorce sloučeny s jinými parametry. Jedná se o 2 případy. V prvním případě byly sloučeny parametry p23 – řízení provozu v době nehody a p24 – místní úprava přednosti v jízdě. Parametr p23 má 4 stavy, přičemž pokud nastane stav 3 – místní úprava, vyplňuje se ještě doplňující parametr p24. V tab. 11 jsou zobrazeny původní stavy, jejich kombinace a stav po sloučení parametrů, kdy vznikl nový parametr – způsob řízení provozu. Pro každý parametr je uvedena hodnota a slovní popis. Stejně jako v případě vyřazených parametrů rozhodnutí o tom, jaké parametry a jakým způsobem budou sloučeny, udělal autor podle svého uvážení a na základě konzultací s vedoucími práce.

tab. 11 - Sloučení parametrů p23 – řízení provozu v době nehody a p24 – místní úprava přednosti v jízdě [zdroj vlastní]

Řízení provozu v době nehody		Místní úprava přednosti v jízdě		Způsob řízení provozu	
Hodnota	Popis	Hodnota	Popis	Hodnota	Popis
1	policistou nebo jiným pověřeným orgánem	-	-	1	policista
2	světelným signalizačním zařízením	-	-	2	SSZ
3	místní úprava	1	světelná signalizace přeprnuta na přerušovanou žlutou	3	dopravní značky
		2	světelná signalizace mimo provoz		
		3	přednost vyznačena dopravními značkami		
		4	přednost vyznačena přenosnými dopravními značkami nebo zařízením		
		5	přednost nevyznačena - vyplývá z pravidel silničního provozu	0	žádný
0	žádný způsob řízení provozu	-	-	0	žádný

Druhým případem byly počty lehce, těžce a smrtelně zraněných při nehodě a celková hmotná škoda. Podle vzorce byla vypočtena celková vážnost nehody (VN), ta byla dále rozdělena na 3 základní stavy: lehká, střední a těžká vážnost nehody. [21]

$$VN = 433 \cdot LZ + 4867,7 \cdot TZ + 19440 \cdot SZ + HS$$

kde VN je vážnost nehody, LZ počet lehce zraněných, TZ počet těžce zraněných, SZ počet smrtelně zraněných a HS hmotná škoda. Veškeré počty platí časový údaj do 24 hodin od nehody.

Pro lehkou nehodu byla autorem dále stanovena horní hranice 50 000 Kč, pro střední nehodu rozpětí 50 001 – 250 000 Kč a pro těžkou nehodu spodní hranice 250 001 Kč. Tyto hranice byly stanoveny autorem po konzultaci s vedoucími práce. Z hlediska celkového rozložení se jedná o zhruba 60 % lehkých nehod, 37 % středně těžkých nehod a 3 % těžkých nehod.

3.2.6 Redukce počtu stavů parametrů

Další úpravou bylo snížení počtu stavů u jednotlivých parametrů. U některých parametrů byly určité stavy zastoupeny velmi málo, proto byly přiřazeny k stavům s podobným významem, případně byly s jinými stavy sloučeny do společného stavu. Dále se také jednalo o parametry, které obsahovaly velké množství stavů, muselo tedy být přistoupeno k redukci z hlediska vhodné interpretace výsledků. I v tomto případě bylo k těmto stavům přistupováno jako ve výše popsaném příkladu. Obdobně jako v předchozích případech se i v případě redukce počtu stavů parametrů jedná o rozhodnutí autora podle svého uvážení a na základě konzultací s vedoucími práce.

V následujících tabulkách budou uvedeny jednotlivé parametry, jejich stavy, úpravy, ke kterým bylo přistoupeno, a jejich odůvodnění. Hodnoty parametrů se v tomto případě změnil z číselného formátu na formát faktorů. Faktory jsou datové objekty, které se používají ke kategorizaci dat a ukládají se jako úrovně (levels). Mohou uchovávat řetězce i čísla (v rámci této práce budou uchovávány pouze řetězce). Jsou užitečné při zpracování dat ve sloupcích, které mají omezený počet unikátních hodnot. Příkladem vhodného využití faktorů může být sloupec s hodnotami Muž/Žena nebo True/False. [22]

K převedení čísel na faktory byla využita funkce *cut*, která převádí vektor čísel na jednotlivé faktory podle zadaných intervalů. Základním vstupem funkce je číselný vektor původních dat *x*, definice intervalů *breaks* (číselný vektor nebo jedno číslo, které udává počet intervalů) a *labels*, které k jednotlivým intervalům přiřazují řetězce nebo čísla označující faktory. Příklad použití bude uveden u popisu úprav parametru Den v týdnu. [23]

Den v týdnu

Den v týdnu byl z důvodu snížení počtu stavů a na základě podobné charakteristiky redukován na 2 stavy – pracovní a víkendový den. Četnosti pro jednotlivé hodnoty a nové zařazení jsou uvedeny v tab. 12. Je patrné, že v pracovní dny došlo k přibližně stejnému počtu nehod, zatímco v sobotu a zejména

v neděli došlo k poklesu počtu. V datech se vyskytovala i hodnota 0, která pravděpodobně reprezentovala neznámý nebo nezjištěný den nehody. Řádky s touto hodnotou byly smazány a dále nevyužity. Celkem se jednalo o 7012 řádků.

Funkce *cut* byla použita v následujícím tvaru:

```
cut(data_celkem$den, breaks = c(0,5,7), labels =
c("pracovní", "víkend"))
```

Argument *breaks* definuje dva intervaly, na které se data budou dělit – (0,5] a (6,7] – a argument *labels* jejich označení – pracovní a víkend.

tab. 12 - Četnosti a nové zařazení hodnot parametru Den v týdnu [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
1	pondělí	37404	pracovní
2	úterý	36685	pracovní
3	středa	38198	pracovní
4	čtvrtek	37887	pracovní
5	pátek	37000	pracovní
6	sobota	21681	víkend
7	neděle	12796	víkend

Čas

Čas nehody byl zaznamenáván ve formátu čtyřmístného čísla HHMM, ze kterého byly získány pouze hodiny. Na základě hodin byly poté stanoveny úrovně faktorů jako denní doby – ráno (6:01-12:00), odpoledne (12:01-18:00), večer (18:01-22:00), noc (22:01-6:00). V datech se objevovala hodnota 2560, pravděpodobně pro neznámý čas nehody, řádky s touto hodnotou byly smazány. Celkem se jednalo o 31750 řádků. Pro přehlednost jsou v tab. 13 uvedeny pouze četnosti pro vytvořené faktory. Vyplývá z ní, že převážná část nehod se stala v odpoledních a večerních hodinách, nejméně potom v ranním intervalu 6-12 hodin.

tab. 13 - Četnosti a nové zařazení hodnot parametru Čas [zdroj vlastní]

Hodnota	Četnost	Nové zařazení
6:01-12:00	15588	ráno
12:01-18:00	70845	odpoledne
18:01-22:00	81035	večer
22:01-6:00	29380	noc

Datum (p2a)

Datum bylo převedeno ve formátu RRRR-MM-DD, který RStudio nerozpoznalo. Pro další práci ale stačilo pracovat pouze s měsíci, které byly získány pomocí funkce *substr*. Ta umožňuje z řetězce část vyjmout a dále s ní pracovat. Měsíce byly následně na základě podobných charakteristik převedeny na roční období, aby došlo ke snížení počtu stavů. V tab. 14 jsou uvedeny četnosti a nové zařazení hodnot. Je patrné, že četnost ve většině měsíců je mezi 18000 a 22000 nehodami, výjimkami jsou únor (nižší počet), říjen (vyšší počet) a prosinec (nižší počet).

tab. 14 - Četnosti a nové zařazení hodnot parametru Datum (p2a) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
1	leden	18981	zima
2	únor	17215	zima
3	březen	19053	jaro
4	duben	19846	jaro
5	květen	21092	jaro
6	červen	21147	léto
7	červenec	18295	léto
8	srpen	19119	léto
9	září	21006	podzim
10	říjen	23061	podzim
11	listopad	18211	podzim
12	prosinec	11637	zima

Lokalita nehody (p5a)

Tento parametr má pouze 2 stavy – intravilán a extravilán, byl tedy ponechán v původním rozdělení. Z tab. 15 je možné vidět, že výrazně převažují nehody v intravilánu, což odpovídá i charakteristice oblastí, ze kterých data pochází.

tab. 15 - Četnosti hodnot parametru Lokalita nehody (p5a) [zdroj vlastní]

Hodnota	Význam	Četnost
1	intravilán	165623
2	extravilán	9694

Druh nehody (p6)

Existuje 10 možných hodnot parametru Druh nehody, jejichž četnosti jsou společně s novým zařazením uvedeny v tab. 16. Sloučení některých stavů bylo provedeno na základě jejich podobné charakteristiky

a nízké četnosti některých stavů. Z tabulky vyplývá, že převažují nehody s jedoucím nebo zaparkovaným vozidlem nebo pevnou překážkou.

tab. 16 - Četnosti a nové zařazení hodnot parametru Druh nehody (p6) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	jiný druh nehody	3146	ostatní
1	srážka s jedoucím nekolejovým vozidlem	106027	s nekolejovým vozidlem
2	srážka s vozidlem zaparkovaným, odstaveným	24494	se zaparkovaným vozidlem
3	srážka s pevnou překážkou	23019	s pevnou překážkou
4	srážka s chodcem	6035	s chodcem
5	srážka s lesní zvěří	4659	se zvířetem
6	srážka s domácím zvířetem	388	se zvířetem
7	srážka s vlakem	54	s kolejovým vozidlem
8	srážka s tramvají	4671	s kolejovým vozidlem
9	havárie	2824	havárie

Druh srážky jedoucích vozidel (p7)

V případě srážky jedoucích vozidel se rozlišuje 5 základních stavů, které jsou uvedeny v tab. 17 společně s četnostmi. Významnou část v tomto případě tvoří situace, kdy se o srážku jedoucích vozidel vůbec nejedná. Pokud ano, pak se nejčastěji jedná o srážku zezadu, nejmenší počet nehod tvořily nehody čelní.

tab. 17 - Četnosti a nové zařazení hodnot parametru Druh srážky jedoucích vozidel (p7) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	nepřichází v úvahu, nejedná se o srážku jedoucích vozidel	64565	srážka není
1	čelní	4744	čelní
2	boční	27477	boční
3	z boku	31880	z boku
4	zezadu	46651	zezadu

Druh pevné překážky (p8)

V případě srážky s pevnou překážkou se rozlišuje 10 základních stavů, které jsou uvedeny v tab. 18 společně s četnostmi. Většinu v tomto případě tvoří situace, kdy se o srážku s pevnou překážkou vůbec nejedná. Pokud ano, pak se nejčastěji jedná o jinou překážku, patník, sloupek (např. dopravní značky) nebo sloup (např. veřejného osvětlení). U hodnot 5 a 8 došlo ke sloučení z důvodu nízkých četností s hodnotou 9 do stavu *ostatní*.

tab. 18 - Četnosti a nové zařazení hodnot parametru Druh pevné překážky (p8) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	nepřichází v úvahu, nejedná se o srážku s pevnou překážkou	152298	není pevná překážka
1	strom	1588	strom
2	sloup (VO, elektrické vedení, ...)	3575	sloup
3	odrazník, patník, sloupek	4508	patník/sloupek/značka
4	svodidlo	3569	svodidlo
5	překážka vzniklá provozem jiného vozidla	69	ostatní
6	zeď, pevná část mostů	2442	zeď
7	závory železničního přejezdu	59	závora přejezdu
8	překážka vzniklá stavební činností	317	ostatní
9	jiná překážka	6892	ostatní

Charakter nehody (p9)

Tento parametr má pouze 2 stavy – nehoda s následky na životě a nehoda pouze s hmotnou škodou, byl tedy ponechán v původním rozdělení. Z tab. 19 je možné vidět, že výrazně převažují nehody pouze s hmotnou škodou, které tvoří zhruba 87 % nehod.

tab. 19 - Četnosti a hodnot parametru Charakter nehody (p9) [zdroj vlastní]

Hodnota	Význam	Četnost
1	s následky na životě	22036
2	pouze s hmotnou škodou	153281

Zavinění nehody (p10)

U zavinění nehody je rozlišováno celkem 8 možných hodnot, přičemž 4 z nich byly sloučeny do společného stavu *ostatní*, zejména kvůli nízkým četnostem, jak vyplývá z tab. 20. Nejčastěji jsou nehody zaviněny řidičem motorového vozidla.

tab. 20 - Četnosti a nové zařazení hodnot parametru Zavinění nehody (p10) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	jiné zavinění	664	ostatní
1	řidičem motorového vozidla	164100	řidič motorové
2	řidičem nemotorového vozidla	1562	řidič nemotorové
3	chodcem	2780	chodec
4	lesní zvěří, domácím zvířectvem	5010	zvíře
5	jiným účastníkem silničního provozu	290	ostatní
6	závadou komunikace	290	ostatní
7	technickou závadou vozidla	621	ostatní

Alkohol u viníka (p11)

V případě parametru alkohol u viníka se kromě stavu, kdy alkohol nebyl zjišťován, rozlišuje i přítomnost drog (a případně jejich kombinace s alkoholem). Samotná přítomnost alkoholu u viníka je rozdělena na 6 úrovní podle obsahu, které ale pro další práci byly sloučeny do jednoho. V převážné většině nehod alkohol v krvi nebyl přítomen, případně nebyl zjišťován, jak vyplývá z tab. 21.

tab. 21 - Četnosti a nové zařazení hodnot parametru Alkohol u viníka (p11) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	nezjišťováno	16740	nezjišťováno
1	ano, obsah alkoholu v krvi do 0,24 ‰	336	ano
2	ne	152735	ne
3	ano, obsah alkoholu v krvi od 0,24 ‰ do 0,5 ‰	490	ano
4	pod vlivem drog	288	drogy
5	pod vlivem alkoholu a drog	65	drogy
6	ano, obsah alkoholu v krvi od 0,5 ‰ do 0,8 ‰	387	ano
7	ano, obsah alkoholu v krvi od 0,8 ‰ do 1,0 ‰	266	ano
8	ano, obsah alkoholu v krvi od 1,0 ‰ do 1,5 ‰	806	ano
9	ano, obsah alkoholu v krvi 1,5 ‰ a více	3204	ano

Hlavní příčina (p12)

U hlavní příčiny nehody se rozlišuje 6 základních kategorií, které poté můžou být dále upřesněny. Pro potřeby práce byly dále ponechány pouze tyto základní kategorie, jak je uvedeno i v tab. 22. Z ní plyne, že nejčastější příčinou nehod je nesprávný způsob jízdy, do kterého spadá mj. nedodržení bezpečné vzdálenosti za vozidlem, nezvládnutí řízení vozidla nebo že se řidič plně nevěnoval řízení vozidla.

tab. 22 - Četnosti hodnot parametru Hlavní příčina (p12) [zdroj vlastní]

Hodnota	Význam	Četnost
100	nezaviněná řidičem	9034
201-209	nepřiměřená rychlost jízdy	14338
301-311	nesprávné předjíždění	1203
401-414	nedání přednosti v jízdě	49496
501-516	nesprávný způsob jízdy	100625
601-615	technická závada vozidla	621

Druh povrchu vozovky (p15)

Co se týče druhu povrchu vozovky, výrazně převažuje živичný povrch, následují dlažba a beton, zbylé 4 zaznamenávané druhy byly kvůli svým nízkým četnostem sloučeny do stavu *ostatní*, jak je patrné z tab. 23.

tab. 23 - Četnosti a nové zařazení hodnot parametru Druh povrchu vozovky (p15) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	nezjištěno	330	ostatní
1	dlažba	12711	dlažba
2	živice	158837	živice
3	beton	2469	beton
4	panely	224	ostatní
5	štěrk	486	ostatní
6	jiný nezpevněný povrch	260	ostatní

Stav povrchu vozovky (p16)

Stav povrchu vozovky je možné zaznamenat 10 různými způsoby, z nichž většina byla kvůli podobnosti nebo nízké četnosti sloučena, jak vyplývá z tab. 24. Z té také vyplývá, že povrch vozovky byl při většině nehod suchý (zhruba 79 %), v dalších asi 18 % se jednalo o mokrý povrch.

tab. 24 - Četnosti a nové zařazení hodnot parametru Stav povrchu vozovky (p16) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	jiný stav povrchu vozovky v době nehody	103	ostatní
1	povrch suchý neznečištěný	138362	suchý
2	povrch suchý znečištěný	340	suchý
3	povrch mokrý	32477	mokrý
4	na vozovce je bláto	67	mokrý
5	na vozovce je náledí, ujetý sníh – posypané	1282	sníh/náledí
6	na vozovce je náledí, ujetý sníh – neposypané	1962	sníh/náledí
7	na vozovce je rozlitý olej, nafta apod.	44	ostatní
8	souvislá sněhová vrstva, rozbředlý sníh	641	sníh/náledí
9	náhla změna stavu vozovky (námraza, místní náledí)	39	sníh/náledí

Stav komunikace (p17)

Stavem komunikace je myšlen jak jeho dlouhodobý stav (např. podélné a příčné vedení), tak i stav aktuální (stav dopravního značení, uzavírky, výtluky apod.). Vzhledem k výrazné převaze dobrého stavu komunikace (hodnota 1) byly všechny ostatní hodnoty sloučeny, jak je uvedeno v tab. 25.

tab. 25 - Četnosti a nové zařazení hodnot parametru Stav komunikace (p17) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
1	dobrý, bez závad	173122	dobrý
2	podélný sklon vyšší než 8 %	207	špatný
3	nesprávně umístěná, znečištěná, chybějící dopravní značka	242	špatný
4	zvlněný povrch v podélném směru	70	špatný
5	souvislé výtluky	198	špatný
6	nesouvislé výtluky	397	špatný
7	trvalé zúžení vozovky	17	špatný
8	příčná stružka, hrbol, vystouplé, propadlé kolejnice	44	špatný
9	neoznačená nebo nedostatečně označená překážka na komunikaci	60	špatný
10	přechodná uzavírka jednoho jízdního pruhu	58	špatný
11	přechodná uzavírka komunikace nebo jízdního pásu	41	špatný
12	jiný (neuvedený) stav nebo závada komunikace	861	špatný

Povětrnostní podmínky (p18)

Povětrnostní podmínky jsou zaznamenávány 8 různými hodnotami, z nichž jednoznačně nejčastější jsou neztížené podmínky následované deštěm. Pro zjednodušení byly podle podobnosti nebo kvůli nízké četnosti určité hodnoty sloučeny, jak je uvedeno v tab. 26.

tab. 26 - Četnosti a nové zařazení hodnot parametru Povětrnostní podmínky (p18) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	jiné ztížené	275	jinak ztížené
1	neztížené	160226	neztížené
2	mlha	521	jinak ztížené
3	na počátku deště, slabý déšť, mrholení apod.	3597	déšť
4	déšť	7923	déšť
5	sněžení	1906	sněžení/náledí
6	tvoří se námraza, náledí	787	sněžení/náledí
7	nárazový vítr (boční, vichřice apod.)	82	jinak ztížené

Viditelnost (p19)

U viditelnosti se rozlišuje 7 hodnot, které se rozdělují podle denní doby a toho, zda byla viditelnost zhoršená, nebo ne. V noci se ještě rozlišuje přítomnost veřejného osvětlení (VO). Z toho autor vycházel při sloučení původních stavů na nové, které jsou 4, jak plyne z tab. 27. Nejčastěji k nehodám z tohoto hlediska docházelo ve dne za nezhoršené viditelnosti, poté v noci s VO za nezhoršené viditelnosti.

tab. 27 - Četnosti a nové zařazení hodnot parametru Viditelnost (p19) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
1	ve dne, viditelnost nezhoršená vlivem povětrnostních podmínek	129903	den nezhoršená
2	ve dne, zhoršená viditelnost (svítání, soumrak)	5416	den zhoršená
3	ve dne, zhoršená viditelnost vlivem povětrnostních podmínek (mlha, sněžení, déšť apod.)	6731	den zhoršená
4	v noci, s veřejným osvětlením, viditelnost nezhoršená vlivem povětrnostních podmínek	26488	noc s VO
5	v noci, s veřejným osvětlením, zhoršená viditelnost vlivem povětrnostních podmínek (mlha, déšť, sněžení apod.)	2397	noc s VO
6	v noci, bez veřejného osvětlení, viditelnost nezhoršená vlivem povětrnostních podmínek	3930	noc bez VO
7	v noci, bez veřejného osvětlení, viditelnost zhoršená vlivem povětrnostních podmínek (mlha, déšť, sněžení apod.)	452	noc bez VO

Rozhledové poměry (p20)

Co se týče rozhledových poměrů, rozlišuje se 7 hodnot, 1 pro dobré poměry a ostatních 6 pro různé důvody špatných poměrů. Vzhledem k výrazné převaze dobrých rozhledových poměrů (asi 99 %) byly všechny ostatní hodnoty sloučeny, jak je uvedeno v tab. 28.

tab. 28 - Četnosti a nové zařazení hodnot parametru Rozhledové poměry (p20) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	jiné špatné	164	špatný
1	dobré	174463	dobrý
2	špatné vlivem okolní zástavby	210	špatný
3	špatné vlivem průběhu komunikace	220	špatný
4	špatné vlivem vegetace – trvale	76	špatný
5	špatné vlivem vegetace – přechodně	11	špatný
6	výhled zakryt stojícím vozidlem	173	špatný

Dělení komunikace (p21)

Parametr dělení komunikace rozlišuje počet jízdních pruhů komunikace a případné směrové rozdělení. Jak plyne z tab. 29, rozlišuje 7 hodnot, z nichž byly sloučeny pouze 2 týkající se čtyřpruhové komunikace. Nejčastěji k nehodám docházelo na dvoupruhových komunikacích, případně na typu komunikace, který nespadá pod dané hodnoty.

tab. 29 - Četnosti a nové zařazení hodnot parametru Dělení komunikace (p21) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	žádná z uvedených	54378	ostatní
1	dvoupruhová	65377	dvoupruhová
2	třípruhová	25463	třípruhová
3	čtyřpruhová s dělicím pásem	12571	čtyřpruhová
4	čtyřpruhová s dělicí čarou	3193	čtyřpruhová
5	vícepruhová	11132	vícepruhová
6	rychlostní komunikace	3203	rychlostní

Situování nehody na komunikaci (p22)

Parametr popisuje, kde na komunikaci k nehodě došlo, nejčastěji se jedná o nehody na jízdním pruhu. Některé hodnoty byly z důvodu svých nízkých četností a na základě podobnosti sloučeny, jak je zobrazeno v tab. 30.

tab. 30 - Četnosti a nové zařazení hodnot parametru Situování nehody na komunikaci (p22) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	žádné z uvedených	6174	ostatní
1	na jízdním pruhu	156287	jízdni pruh
2	na odstavném pruhu	699	odstavný pruh/krajnice
3	na krajnici	2445	odstavný pruh/krajnice
4	na odbočovacím, připojovacím pruhu	425	jízdni pruh
5	na pruhu pro pomalá vozidla	93	jízdni pruh
6	na chodníku nebo ostrůvku	829	chodník/cyklostezka
7	na kolejích tramvaje	6320	koleje tramvaje
8	mimo komunikaci	1946	mimo komunikace
9	na stezce pro cyklisty	99	chodník/cyklostezka

Specifická místa a objekty (p27)

Parametr se zabývá specifickými místy a objekty, které se vyskytují v blízkosti zaznamenané nehody. U většiny nehod žádná místa nebo objekty nebyla zaznamenána. Na základě podobnosti byly některé hodnoty sloučeny do nového stavu, jak je uvedeno v tab. 31.

tab. 31 - Četnosti a nové zařazení hodnot parametru Specifická místa a objekty (p27) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	žádné nebo žádné z uvedených	130537	žádné/ostatní
1	přechod pro chodce	10534	přechod pro chodce
2	v blízkosti přechodu pro chodce (do vzdálenosti 20 m)	17348	přechod pro chodce
3	železniční přejezd nezabezpečený závorami ani světelným výstražným zařízením	63	železniční přejezd
4	železniční přejezd zabezpečený	107	železniční přejezd
5	most, nadjezd, podjezd, tunel	2276	most/nadjezd/podjezd/tunel
6	zastávka autobusu, trolejbusu, tramvaje s nástup. ostrůvkem	896	zastávka
7	zastávka tramvaje, autobusu, trolejbusu bez nástup. ostrůvku	822	zastávka
8	výjezd z parkoviště, lesní cesty apod.	311	parkoviště
9	čerpadlo pohonných hmot	1372	čerpadlo PHM
10	parkoviště přiléhající ke komunikaci	11051	parkoviště

Směrové poměry (p28)

Směrové poměry rozlišují, zda se nehoda udála v přímém úseku, zatáčce nebo na různých druzích křižovatek. Ke sloučení hodnot došlo pouze u dvou hodnot. Jak vyplývá z tab. 32, nejčastěji k nehodám docházelo v přímých úsecích, z křižovatek potom na průsečných.

tab. 32 - Četnosti a nové zařazení hodnot parametru Směrové poměry (p28) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
1	přímý úsek	112570	přímý úsek
2	přímý úsek po projetí zatáčkou	3241	přímý úsek
3	zatáčka	9244	zatáčka
4	křižovatka průsečná – čtyřramenná	25993	křižovatka průsečná
5	křižovatka styková – tříramenná	19990	křižovatka styková
6	křižovatka pěti a víceramenná	2591	křižovatka ostatní
7	kruhový objezd	1688	kruhový objezd

Místo nehody (p35)

Hodnoty parametru místo nehody vychází ze způsobu zaznamenávání pohybu vozidel v křižovatce, který vedl k nehodě. Křižovatka je rozdělena na zóny 1-9, které se určují podle tvaru křižovatky, zóny 1-8 označují ramena křižovatky, zóna 9 samotné křížení komunikací. Jak je uvedeno v tab. 33, nejčastějším místem nehody byly úseky mimo křižovatku, další hodnoty byly podle podobnosti sloučeny.

tab. 33 - Četnosti a nové zařazení hodnot parametru Místo nehody (p35) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	mimo křižovatku	125099	mimo křižovatku
10	na křižovatce (místních nebo účelových komunikací nebo mezilehlá)	26228	střed křižovatky
11-18	uvnitř zóny 1-8 předmětné křižovatky	1366	ramena křižovatky
19	na křižovatce	21528	střed křižovatky
22-28	na vjezdové nebo výjezdové části větve při mimoúrovňovém křížení	35	mimo křižovatku
29	mimo zónu 11-19 a 22-28	1061	mimo křižovatku

Druh komunikace (p36)

U parametru druh komunikace se rozlišuje celkem 9 hodnot. Komunikacemi, kde docházelo k nehodám nejčastěji, jsou komunikace sledované ve vybraných městech a místní komunikace, což odpovídá převážně městskému charakteru zpracovávaných oblastí. Na základě podobné charakteristiky došlo ke sloučení 2 hodnot týkajících se účelových komunikací, jak uvádí tab. 34.

tab. 34 - Četnosti a nové zařazení hodnot parametru Druh komunikace (p36) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	dálnice	1222	dálnice
1	silnice 1. třídy	2501	1. třída
2	silnice 2. třídy	5271	2. třída
3	silnice 3. třídy	5734	3. třída
4	uzel (ve vybraných městech)	21194	uzel
5	komunikace sledovaná (ve vybraných městech)	74077	sledovaná město
6	komunikace místní	55300	místní
7	komunikace účelová (polní a lesní cesty)	1468	účelová
8	komunikace účelová (parkoviště, odpočívky)	8550	účelová

Druh vozidla (p44)

Parametr druh vozidla umožňuje rozlišovat až 19 různých druhů vozidla, které nehodu zavinilo. Některé z těchto druhů se ale v celém vzorku nevyskytly ani jednou. Nejčastějším druhem naopak byl osobní automobil. Vzhledem k vysokému počtu hodnot a nízkým četnostem některých z nich došlo k několika sloučením do nových stavů, jak vyplývá z tab. 35.

tab. 35 - Četnosti a nové zařazení hodnot parametru Druh vozidla (p44) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	moped	81	motocykl
1	malý motocykl (do 50 ccm)	215	motocykl
2	motocykl (včetně sidecarů, skútrů apod.)	2506	motocykl
3	osobní automobil bez přívěsu	136177	osobní
4	osobní automobil s přívěsem	806	osobní
5	nákladní automobil (vč. multikáry, autojeřábu, cisterny)	19673	nákladní
6	nákladní automobil s přívěsem	1297	nákladní
7	nákladní automobil s návěsem	5353	nákladní
8	autobus	5269	autobus/trolejbus

9	traktor (i s přívěsem)	97	ostatní
10	tramvaj	1777	tramvaj
11	trolejbus	0	autobus/trolejbus
12	jiné motorové vozidlo	312	ostatní
13	jízdní kolo	1716	kolo
14	povoz, jízda na koni	10	ostatní
15	jiné nemotorové vozidlo	26	ostatní
16	vlak	0	ostatní
17	nezjištěno, řidič ujel	0	ostatní
18	jiný druh vozidla	2	ostatní

Smyk (p49)

Parametr smyk má pouze 2 stavy – ano (ke smyku došlo) a ne (ke smyku nedošlo), byl tedy ponechán v původním rozdělení. Z tab. 36 je možné vidět, že výrazně převažují nehody, při kterých ke smyku nedošlo. Tyto nehody tvoří zhruba 95 % nehod.

tab. 36 - Četnosti hodnot parametru Smyk (p49) [zdroj vlastní]

Hodnota	Význam	Četnost
0	ne	167232
1	ano	8085

Únik hmot (p50b)

U úniku hmot je rozlišováno 5 hodnot. Pokud k úniku došlo, zaznamenává se podrobně, zda se jednalo o látku pohonnou nebo přepravovanou. U přepravované látky se poté ještě rozlišuje skupenství. Vzhledem k výrazné většině nehod, kdy k úniku nedošlo, a k nízkým či nulovým četnostem ostatních hodnot byly zbylé hodnoty sloučeny, jak zobrazuje tab. 37.

tab. 37 - Četnosti a nové zařazení hodnot parametru Únik hmot (p50b) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	žádné z uvedených	170984	ne
1	došlo k úniku pohonných hmot, oleje, chladicího média apod.	4331	ano
2	došlo k úniku jiných nebezpečných látek pevných	0	ano
3	došlo k úniku jiných nebezpečných látek kapalných	2	ano
4	došlo k úniku jiných nebezpečných látek plyných	0	ano

Kategorie řidiče (p55a)

Kategorie řidiče rozlišuje základní skupiny řidičského oprávnění, které bylo zjištěno u viníka nehody a další situace, kdy řidičské oprávnění řidič nemá nebo nebylo možné ho zjistit. Nejčastěji se jednalo o držitele oprávnění skupiny B. Vzhledem k nízkým četnostem některých hodnot byly sloučeny do stavu ostatní, jak je uvedeno v tab. 38.

tab. 38 - Četnosti a nové zařazení hodnot parametru Kategorie řidiče (p55a) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	nezjištěno (např. u cizinců)	285	ostatní
1	s řidičským oprávněním skupiny A	159	ostatní
2	s řidičským oprávněním skupiny B	122952	B
3	s řidičským oprávněním skupiny C	35194	C
4	s řidičským oprávněním skupiny D	11713	D
5	s řidičským oprávněním skupiny T	52	ostatní
6	s řidičským oprávněním skupiny A a s omezením do 50 ccm	49	ostatní
7	bez příslušného řidičského oprávnění	1212	bez oprávnění
8	ostatní řidiči vozidel (cyklisté, strojvedoucí)	3695	ostatní
9	nezjištěno, řidič místo nehody opustil	6	ostatní

Nejvyšší ukončené vzdělání (p55b)

Nejvyšší ukončené vzdělání je zaznamenáváno v 5 různých hodnotách, které byly ponechány beze změny, jak vyplývá z tab. 39. Pokud bylo nejvyšší ukončené vzdělání zjištěno, nejčastěji se jednalo o vzdělání středoškolské.

tab. 39 - Četnosti hodnot parametru Nejvyšší ukončené vzdělání (p55b) [zdroj vlastní]

Hodnota	Význam	Četnost
0	nezjištěno	88280
1	základní	5617
2	učňovské	4433
3	středoškolské	62810
4	vysokoškolské	14177

Délka řídicí praxe v řízení vozidla (p56)

Délka řídicí praxe byla zaznamenávána dvojcíslím, přičemž hodnoty se pohybovaly v rozmezí 0-70 let. Pro zjednodušení byly navrženy intervaly, které jsou společně s jejich četnostmi zobrazeny v tab. 40. V datovém souboru se často vyskytovala hodnota XX, která byla nahrazena hodnotou *nezjištěno*. Ze zjištěných intervalů nejvíce řidičů řídilo mezi 6-10 a 11-20 lety.

tab. 40 - Četnosti hodnot parametru Délka řídicí praxe v řízení vozidla (p56) [zdroj vlastní]

Hodnota	Četnost
nezjištěno	70842
<2	16248
3-5	17702
6-10	25211
11-20	25108
21-50	19580
>50	626

Stav řidiče (p57)

Stavem řidiče je myšlen jeho stav v době nehody. Ten může být ovlivněn fyzickou indispozicí nebo omamnými látkami, jak je i naznačeno v nově zařazených hodnotách. Nejčastěji byl ale zaznamenán dobrý stav řidiče, jak plyne z tab. 41.

tab. 41 - Četnosti a nové zařazení hodnot parametru Stav řidiče (p57) [zdroj vlastní]

Hodnota	Význam	Četnost	Nové zařazení
0	jiný nepříznivý stav	565	ostatní
1	dobrý	168498	dobrý
2	unaven, usnul, náhlá fyzická indispozice	509	fyzická indispozice
3	pod vlivem léků, narkotik	327	pod vlivem omamných látek
4	pod vlivem alkoholu, obsah v krvi do 0,99 ‰	1394	pod vlivem omamných látek
5	pod vlivem, obsah v krvi 1,00 ‰ a více	3802	pod vlivem omamných látek
6	nemoc, úraz apod.	135	fyzická indispozice
7	invalida	31	fyzická indispozice
8	řidič při jízdě zemřel (infarkt apod.)	49	fyzická indispozice
9	pokus o sebevraždu, sebevražda	7	ostatní

Vnější ovlivnění řidiče (p58)

Co se týče vnějšího ovlivnění řidiče, rozlišuje se 6 hodnot, 1 pro žádné ovlivnění a zbylých 5 pro různé důvody vnějšího ovlivnění. Vzhledem k výrazné převaze nehod bez vnějšího ovlivnění (asi 98 %) byly všechny ostatní hodnoty sloučeny, jak je uvedeno v tab. 42.

tab. 42 - Četnosti a nové zařazení hodnot parametru Vnější ovlivnění řidiče (p58) [zdroj vlastní]

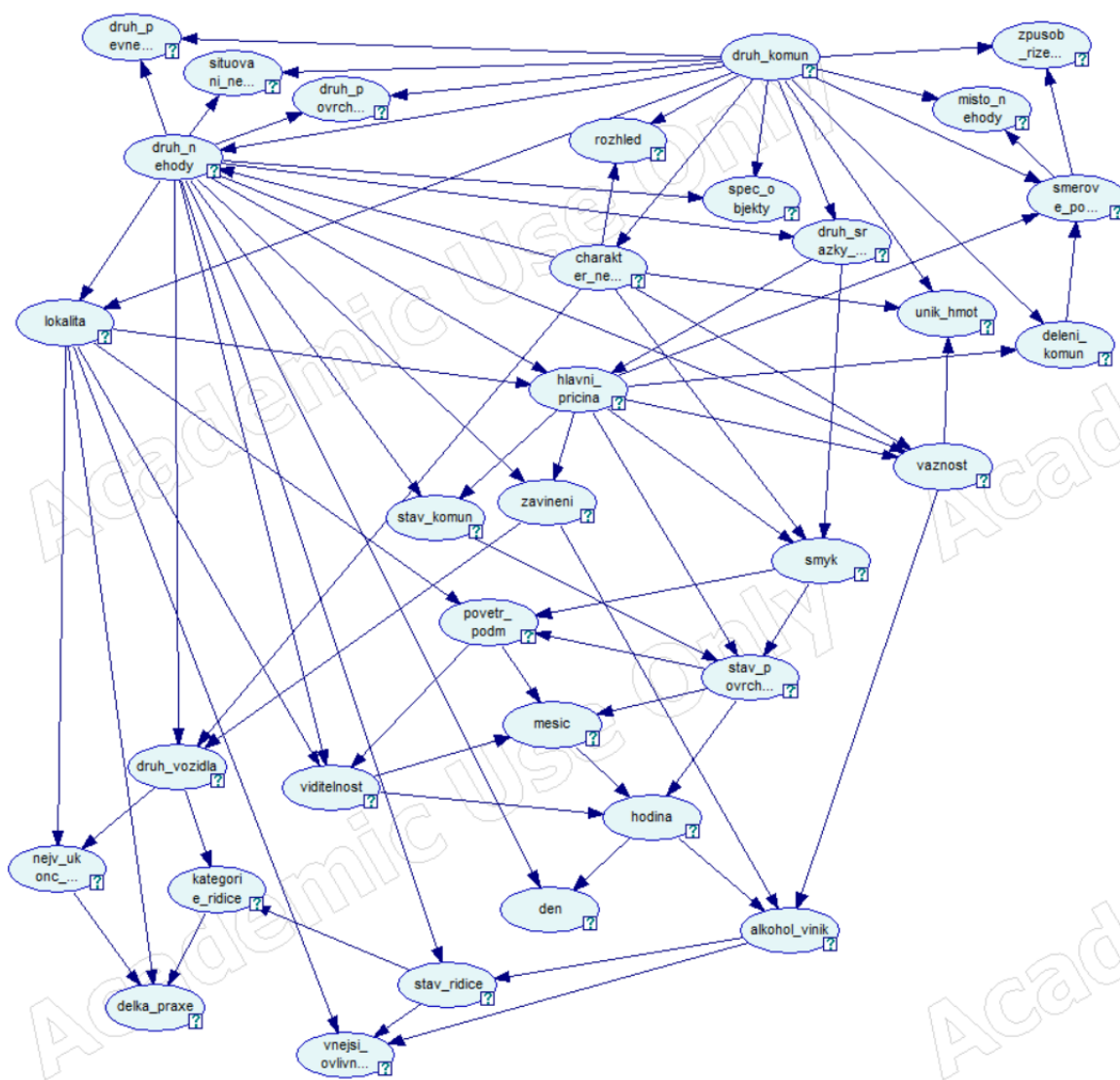
Hodnota	Význam	Četnost	Nové zařazení
0	jiné ovlivnění	1815	ano
1	řidič nebyl ovlivněn	172575	ne
2	oslněn sluncem	226	ano
3	oslněn světlomety jiného vozidla	29	ano
4	ovlivněn jednáním jiného účastníka silničního provozu	512	ano
5	ovlivněn při vyhýbání lesní zvěří, domácím zvířectvu apod.	160	ano

3.3 Výstupy

V této kapitole je popsána tvorba sítě z připravených dat. Dále je vzniklá síť popsána z hlediska struktury a v závěru kapitoly je popsán hlavní výstup práce. Jde o tabulku, která zobrazuje největší změnu v pravděpodobnosti při zavedení evidence *těžká* do uzlu *vážnost nehody*. Tato tabulka zahrnuje 10 stavů, u kterých došlo k největší změně při vyšší vážnosti nehody.

3.3.1 Popis vytvořené Bayesovské sítě

Pomocí hill-climbing algoritmu popsaného v kapitole 2.3.4 byla na základě připraveného datového souboru vytvořena Bayesovská síť. Tvoří ji 33 uzlů a celkem 77 vazeb mezi nimi. Vytvořená Bayesovská síť je zobrazena na obr. 15, součástí příloh je i výsledná síť ve formátu .xdsl používaném softwarem GeNIe jako příloha 2.



obr. 15 - Vytvořená Bayesovská síť [zdroj vlastní]

V tab. 43 jsou pro lepší přehlednost uvedeni rodiče a děti každého uzlu vytvořené Bayesovské sítě. V síti existuje 1 uzel, který nemá žádné rodiče a pouze děti, tj. všechny vazby v něm začínají, ale žádná nekončí. Jedná se o uzel *druh_komun*, který má celkem 14 dětí. Naopak uzlů, které mají pouze rodiče a žádné děti, tj. všechny vazby v nich pouze končí, je ve vytvořené síti celkem 11. Jedná se o uzly *den*, *druh_pevne_prekazky*, *druh_povrchu_vozovky*, *rozhled*, *situovani_nehody*, *spec_objekty*, *misto_nehody*, *unik_hmot*, *delka_praxe*, *vnejsi_ovlivneni_ridice*, a *způsob_rizeni_provozu*. Z hlediska celkového počtu vazeb patří mezi nejvýznamnější uzly *druh_nehody* (2 rodiče + 14 dětí = 16 vazeb), *druh_komun* (0 R + 14 D = 14 vazeb), *hlavni_pricina* (3 R + 7 D = 10 vazeb), *lokalita* (2 R + 6 D = 8 vazeb) a *charakter_nehody* (1 R + 6 D = 7 vazeb). Zbylé uzly mají zpravidla celkem 2-5 vazeb, žádný z uzlů nemá pouze jednu vazbu.

tab. 43 - Rodiče a děti uzlů vytvořené Bayesovské sítě [zdroj vlastní]

Uzel	Rodiče	Děti
<i>druh_komun</i>	-	<i>lokalita</i> , <i>druh_nehody</i> , <i>druh_srazky_vozidel</i> , <i>druh_pevne_prekazky</i> , <i>charakter_nehody</i> , <i>druh_povrchu_vozovky</i> , <i>rozhled</i> , <i>deleni_komun</i> , <i>situovani_nehody</i> , <i>spec_objekty</i> , <i>smerove_pomery</i> , <i>misto_nehody</i> , <i>unik_hmot</i> , <i>zpusob_rizeni_provozu</i>
<i>den</i>	<i>druh_nehody</i> , <i>hodina</i>	-
<i>lokalita</i>	<i>druh_komun</i> , <i>druh_nehody</i>	<i>hlavni_pricina</i> , <i>povetr_podm</i> , <i>viditelnost</i> , <i>nejv_ukonc_vzdelani</i> , <i>delka_praxe</i> , <i>vnejsi_ovlivneni_ridice</i>
<i>druh_nehody</i>	<i>druh_komun</i> , <i>charakter_nehody</i>	<i>den</i> , <i>lokalita</i> , <i>druh_srazky_vozidel</i> , <i>druh_pevne_prekazky</i> , <i>zavineni</i> , <i>hlavni_pricina</i> , <i>druh_povrchu_vozovky</i> , <i>stav_komun</i> , <i>viditelnost</i> , <i>situovani_nehody</i> , <i>spec_objekty</i> , <i>druh_vozidla</i> , <i>stav_ridice</i> , <i>vaznost</i>

druh_srazky_vozidel	druh_komun, druh_nehody	hlavni_pricina, smyk
druh_pevne_prekazky	druh_komun, druh_nehody	-
charakter_nehody	druh_komun	druh_nehody, rozhled, druh_vozidla, smyk, unik_hmot, vaznost
zavineni	druh_nehody, hlavni_pricina	alkohol_vinik, druh_vozidla
alkohol_vinik	zavineni, vaznost, hodina	stav_ridice, vnejsi_ovlivneni_ridice
hlavni_pricina	lokalita, druh_nehody, druh_srazky_vozidel	zavineni, stav_povrchu_vozovky, stav_komun, deleni_komun, smerove_pomery, smyk, vaznost
druh_povrchu_vozovky	druh_komun, druh_nehody	-
stav_povrchu_vozovky	hlavni_pricina, stav_komun, smyk	povetr_podm, mesic, hodina
stav_komun	druh_nehody, hlavni_pricina	stav_povrchu_vozovky
povetr_podm	lokalita, stav_povrchu_vozovky, smyk	viditelnost, mesic
viditelnost	lokalita, druh_nehody, povetr_podm	mesic, hodina
rozhled	druh_komun, charakter_nehody	-
deleni_komun	druh_komun, hlavni_pricina	smerove_pomery
situovani_nehody	druh_komun, druh_nehody	-
spec_objekty	druh_komun, druh_nehody	-
smerove_pomery	druh_komun, hlavni_pricina, deleni_komun	misto_nehody, zpusob_rizeni_provozu
misto_nehody	druh_komun, smerove_pomery	-
druh_vozidla	druh_nehody, charakter_nehody, zavineni	kategorie_ridice, nejv_ukonc_vzdelani
smyk	druh_srazky_vozidel, charakter_nehody, hlavni_pricina	stav_povrchu_vozovky, povetr_podm,
unik_hmot	druh_komun, charakter_nehody, smyk, vaznost	-
kategorie_ridice	druh_vozidla, stav_ridice	delka_praxe
nejv_ukonc_vzdelani	lokalita, druh_vozidla	delka_praxe
delka_praxe	lokalita, kategorie_ridice, nejv_ukonc_vzdelani	-

stav_ridice	druh_nehody, alkohol_vinik	kategorie_ridice, vnejsi_ovlivneni_ridice
vnejsi_ovlivneni_ridice	lokalita, alkohol_vinik, stav_ridice	-
vaznost	druh_nehody, charakter_nehody, hlavni_pricina	alkohol_vinik, unik_hmot
mesic	stav_povrchu_vozovky, povetr_podm, viditelnost	hodina
hodina	stav_povrchu_vozovky, viditelnost, mesic	den, alkohol_vinik
zpusob_rizeni_provozu	druh_komun, smerove_pomery	-

3.3.2 Popis zjištěných výstupů

Cílem práce je ukázat, které stavy kterých uzlů se nejvíce změní při zavedení evidence *těžká* do uzlu vážnost. Porovnávány byly všechny hodnoty ve všech uzlech sítě před zavedením evidence a po jejím zavedení. Jedná se tedy o rozdíl mezi marginální a podmíněnou pravděpodobností stavů. V tab. 44 je zobrazeno 10 uzlů s největší kladnou změnou, jde tedy o stavy, které jsou se zvýšením vážnosti nehody spojeny nejvíce. Tabulka se změnami všech stavů, kterých je celkem 144, je součástí příloh jako příloha 3 ve formátu .xlsx.

Fakt, že stav s největší kladnou změnou je právě stav s následky na zdraví nebo životě uzlu charakter nehody, lze označit za ověření správného sestavení a výpočtu sítě. Jedná se totiž o uzel, který má pouze 2 stavy – nehoda pouze s hmotnou škodou, nebo nehoda s následky na zdraví nebo životě – a vzhledem k tomu, že hodnota vážnosti nehody byla vypočítána z počtu lehce, těžce a smrtelně zraněných a hmotné škody, bylo možné očekávat, že právě nehody s následky na zdraví nebo životě budou významně souviset s vyšší vážností nehody. Jedná se o nárůst o zhruba 20 % po zavedení evidence.

Mezi další stavy, které bylo možné do jisté míry očekávat na vyšších místech, patří druh nehody – s kolejovým vozidlem a situování nehody – koleje tramvaje. K tomu vedla úvaha, že v kolejových vozidlech (v tomto případě se jedná o vlaky a tramvaje) obecně cestuje více lidí, zvyšuje se tedy pravděpodobnost většího počtu zraněných i hmotné škody. Dalším takovým stavem je i únik hmot, který výrazně zvyšuje hmotnou škodu nehody.

Naopak méně očekávané stavy, které se objevily takto vysoko, jsou druh nehody – s nekolejovým vozidlem. Tento stav autor očekával alespoň níže než nehody s kolejovým vozidlem, kvůli výše uvedeným důvodům, ale nakonec došlo k vyššímu nárůstu o 9 % právě u tohoto stavu. K tomu došlo zřejmě na úkor ostatních stavů, jako srážka se zaparkovaným vozidlem, chodcem nebo zvířetem. Zajímavými jsou také 2 stavy uzlu hlavní příčina. Jedná se o stavy nedání přednosti a nepřiměřená rychlost, které se zvýšily o 8,4 % a 6,6 %. Zejména u nedání přednosti lze najít souvislost s dvěma druhy srážky vozidel, které se v tabulce vyskytují. Jedná se totiž o srážky boční a z boku, které jsou typické pro nedání přednosti v křižovatce. Zároveň jsou také častější v městském provozu, který v datovém souboru převažuje.

tab. 44 - Stavy s největší kladnou změnou po zavedení evidence [zdroj vlastní]

Č.	Uzel	Stav	Marg. [%]	Cond. [%]	Změna [%]
1	charakter_nehody	s nasledky na zdravi nebo zivote	12,571	32,777	20,206
2	druh_nehody	s nekolejovym vozidlem	60,482	69,559	9,077
3	unik_hmot	ano	2,426	11,448	9,022
4	hlavni_pricina	nedani prednosti	28,245	36,623	8,378
5	hlavni_pricina	neprimerena rychlost	8,180	14,734	6,554
6	druh_srazky_vozidel	z boku	18,186	24,706	6,520
7	druh_nehody	s kolejovym vozidlem	2,696	8,827	6,131
8	druh_srazky_vozidel	bocni	15,673	21,125	5,452
9	misto_nehody	stred krizovatky	27,721	33,088	5,367
10	situovani_nehody	koleje tramvaje	3,605	8,808	5,203

Na opačné straně se nachází stavy, u kterých došlo po zavedení evidence těžká do uzlu vážnost k největšímu poklesu. Jde tedy o stavy, u kterých se pravděpodobnost výskytu při vážné nehodě nejvíce snížila. V tab. 45 je zobrazeno 10 stavů s největším poklesem pravděpodobnosti po zavedení evidence. Největší pokles zaznamenal stav nehoda pouze s hmotnou škodou v uzlu charakter nehody. Tato skutečnost logicky plyne z toho, že uzel má pouze 2 stavy, nárůst jednoho stavu musí tedy odpovídat poklesu stavu druhého. Tyto stavy navíc svým rozdělením úzce souvisí s výpočtem vážnosti nehody, jak bylo již naznačeno v předchozím odstavci. Podobnou logiku lze použít i pro únik hmot, jen v této tabulce se jedná o ty situace, kdy k němu nedošlo.

Mezi stavy, které v tabulce bylo možné očekávat, patří stav, kdy se nejedná o srážku jedoucích vozidel, jelikož je pravděpodobnější nižší počet zraněných i hmotná škoda, nehody, ke kterým došlo v přímém úseku (oproti zatáčce nebo křižovatkám), nebo nehody na místních komunikacích, pravděpodobně díky nižším rychlostem oproti dálnicím nebo silnicím 1. třídy.

Zajímavým stavem, který se v tabulce objevuje, je způsob řízení provozu – žádný. Pokles tohoto stavu mírně odporuje poznatkům, že křižovatky řízené SSZ jsou obecně bezpečnější.

tab. 45 - Stavy s největší zápornou změnou po zavedení evidence [zdroj vlastní]

Č.	Uzel	Stav	Marg. [%]	Cond. [%]	Změna [%]
1	charakter_nehody	pouze hmotna skoda	87,429	67,223	-20,206
2	druh_srazky_vozidel	srazka neni	36,822	21,614	-15,209
3	hlavni_pricina	nespravny zpusob jizdy	57,387	46,226	-11,161
4	unik_hmot	ne	97,574	88,552	-9,022
5	druh_nehody	se zaparkovanim vozidlem	13,971	6,064	-7,906
6	smerove_pomery	primy usek	65,584	58,201	-7,383
7	druh_komun	mistni	31,542	25,111	-6,431
8	misto_nehody	mimo krizovatku	71,489	65,851	-5,638
9	zpusob_rizeni_provozu	zadny	71,112	65,517	-5,595
10	alkohol_vinik	nezjistovano	9,550	4,270	-5,281

V tab. 46 jsou zobrazeny stavy, u kterých došlo k největší absolutní změně. V jedné tabulce jsou tedy stavy jak s kladnou, tak se zápornou změnou po zavedení evidence. To, jestli došlo k nárůstu (+), nebo poklesu (-), je uvedeno v posledním sloupci.

tab. 46 - Stavy s největší absolutní změnou po zavedení evidence [zdroj vlastní]

Č.	Uzel	Stav	Abs. změna [%]	+/-
1	charakter_nehody	s nasledky na zdravi nebo zivote	20,206	+
2	charakter_nehody	pouze hmotna skoda	20,206	-
3	druh_srazky_vozidel	srazka neni	15,209	-
4	hlavni_pricina	nespravny zpusob jizdy	11,161	-
5	druh_nehody	s nekolejovym vozidlem	9,077	+
6	unik_hmot	ne	9,022	-
7	unik_hmot	ano	9,022	+
8	hlavni_pricina	nedani prednosti	8,378	+
9	druh_nehody	se zaparkovanim vozidlem	7,906	-
10	smerove_pomery	primy usek	7,383	-
11	hlavni_pricina	neprimerena rychlost	6,554	+
12	druh_srazky_vozidel	z boku	6,520	+
13	druh_komun	mistni	6,431	-
14	druh_nehody	s kolejovym vozidlem	6,131	+
15	misto_nehody	mimo krizovatku	5,638	-
16	zpusob_rizeni_provozu	zadny	5,595	-
17	druh_srazky_vozidel	bocni	5,452	+
18	misto_nehody	stred krizovatky	5,367	+
19	alkohol_vinik	nezjistovano	5,281	-
20	situovani_nehody	koleje tramvaje	5,203	+

4 Závěr

Cílem diplomové práce bylo na základě zpracování nehodových dat a vytvoření Bayesovské sítě identifikovat stavy jednotlivých parametrů, které se nejvíce pojí se zvýšenou vážností nehody. Tyto stavy byly zjištěny porovnáním marginální pravděpodobnosti, kdy ve vytvořené Bayesovské síti nebyla zavedena žádná evidence, a podmíněné pravděpodobnosti při zavedení evidence *těžká* do uzlu *vážnost nehody*. Na základě jejich rozdílu byly zjištěny stavy s největší kladnou změnou, u kterých je vyšší pravděpodobnost, že nastanou u nehod s vyšší vážností. Dále byly v práci uvedeny i stavy s největší zápornou změnou, tedy stavy, u kterých se pravděpodobnost výskytu u vážných nehod nejvíce snížila, a také přehled stavů s největší absolutní změnou.

Vytvoření sítě předcházelo zpracování 2 datových souborů, které obsahovaly data o nehodách z let 2011-2021 na území hl. m. Prahy a přilehlých okresů Praha-východ a Praha-západ. Nejdříve byl popsán soubor z hlediska zpracovávaných oblastí a způsobu, jakým se data zaznamenávají. Poté se jednalo především o výběr relevantních parametrů, které byly do Bayesovské sítě zařazeny, případně vyřazení parametrů, které pro nebyly vhodné nebo relevantní. Dále to také byly úpravy jednotlivých stavů, především jejich slučování, které bylo provedeno kvůli snížení výpočetní náročnosti a lepší interpretovatelnosti. Stavy byly slučovány zejména na základě podobné charakteristiky, případně kvůli nízké četnosti.

Práce také dokázala, že Bayesovské sítě patří mezi užitečné nástroje pro analýzu nehodových dat, zejména pro komplexní a složité vztahy mezi jednotlivými parametry a práci s neurčitostí za zachování přijatelné výpočetní náročnosti. Příkladem může být srovnání s diplomovou prací zabývající se analýzou diskretních nehodových dat. [24] Ta analyzuje data ze stejného zdroje, tedy formulářů Policie ČR, pomocí histogramů, které zobrazují četnosti kombinací stavů několika parametrů. Sama autorka zmiňuje, že kvůli vysokému počtu možných kombinací musela přistoupit k redukci parametrů, ale následně i stavů u těch parametrů, které vybrala. V této práci autor také přistoupil k redukci stavů, ale v mnohem menším množství. Bayesovské sítě tedy tuto nevýhodu eliminovaly a zvládly práci i s velkým počtem stavů. V českém akademickém prostředí se využitím Bayesovských sítí pro analýzu konkrétních nehodových dat dosud nikdo nezabýval, lze ale najít obdobně zaměřené práce v zahraničí. Jedná se o práce, které jsou v seznamu zdrojů uvedeny jako [12], [25], [26] a [27]. Všechny tyto práce se zabývaly využitím Bayesovských sítí pro analýzu příčin nehod ve spojitosti s jejich vážností v Číně, Etiopii, respektive Španělsku. Vzhledem k rozdílné metodice zaznamenávání údajů o nehodách v různých zemích a také jinému postupu autorů při přípravě a zpracování dat ale nelze dost dobře porovnat výstupy uvedených prací s touto. Některé práce použily Bayesovské sítě k sestavení modelu predikce nehod nebo se zaměřily na analýzu vážnosti nehody za určité situace (například v případě zdroje [27] pouze na dvoupruhové dálnici), čímž se od této práce a jejích výstupů odlišují.

V českém prostředí se analýza nehod zaměřuje spíše na identifikaci kritických míst na komunikacích. K tomu slouží i metodika vytvořená CDV, která se zabývá multifaktorovou analýzou dopravní nehodovosti. Popisuje využití statistických metod, jako je korelace, logistická regrese a zobecněný lineární model, k identifikaci rizikových faktorů v křižovatkách a v mezikřižovatkových úsecích. K analýze využívá kromě dat Policie ČR i další zdroje, např. Celostátní sčítání dopravy nebo Silniční databanku ŘSD ČR. Jejím výstupem jsou faktory (vztaženo k této práci faktory nejlépe odpovídají uzlům sítě), které jsou nejvíce rizikové pro vznik nehody. Analýza ale nejde až na úroveň jednotlivých stavů těchto faktorů. [28]

Mezi okolnosti, které mohly ovlivnit výstupy práce, lze jmenovat zejména závislost na subjektivním záznamu některých parametrů nehody prostřednictvím policisty, který vyplňuje příslušný formulář, jak je zmíněno podkapitole 3.2.4. Subjektivita v záznamu parametrů by následně ovlivnila i Bayesovskou síť z hlediska struktury a lokálních distribucí vypočtených algoritmem, jelikož by se jednalo o jiná vstupní data.

Je nutné také zmínit vklad autora, který prováděl různé operace s parametry a jejich stavy, což bezesporu mělo vliv na vytvořenou Bayesovskou síť i výstupy z ní. Jiná osoba by mohla dojít k jiným úpravám, což by samozřejmě vedlo i k jiným výstupům. I proto byl ale autorův postup průběžně konzultován, aby byla zajištěna smysluplnost postupu a co nejkvalitnější příprava dat pro tvorbu sítě. Dále je podle autora možné na tuto práci navázat zaměřením se na jiné oblasti než ty, které vybral pro zpracování autor, nebo zaváděním evidence do jiných uzlů a jejich stavů. Lze předpokládat, že při výběru oblasti, kde je jiný charakter provozu, převažují jiné druhy komunikací i jiná skladba dopravy, se jako kritické projeví zcela jiné stavy a jiné parametry. Bylo by zajímavé tyto změny porovnat, ale také naopak nacházet společné body, které výběr oblasti neovlivnil a vyskytly se ve více případech. Možností z tohoto hlediska je se například zaměřit na porovnání podle charakteru dopravního provozu (dálnice, hl. m. Praha, ostatní města, vesnice, úseky mezi obcemi). Konkrétní uzly, které by bylo podle autora zajímavé dále analyzovat, jsou mj. druh komunikace (dálnice, I., II. a III. třída, místní a účelové komunikace) nebo ty týkající se osoby řidiče (délka řidičské praxe, vzdělání, věk).

Seznam zdrojů

- [1] Road traffic injuries. World Health Organization [online]. [cit. 2023-05-03]. Dostupné z: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>
- [2] Dopravní nehody v ČR. *Centrum dopravního výzkumu, v. v. i.* [online]. [cit. 2023-05-03]. Dostupné z: <https://nehody.cdv.cz/>
- [3] VOLEK, Josef a Bohdan LINDA, 2012. *Teorie grafů - aplikace v dopravě a veřejné správě*. Pardubice: Univerzita Pardubice. ISBN ISBN 978-80-7395-225-9.
- [4] MRKVIČKA, Tomáš, 2016. *Základy teorie pravděpodobnosti a matematické statistiky* [online]. České Budějovice: Jihočeská univerzita [cit. 2023-05-02]. Dostupné z: <http://home.ef.jcu.cz/~mrkvicka/vyuka/TPMSA.pdf>
- [5] *Wikipedie: Otevřená encyklopedie: Teorie pravděpodobnosti* [online]. c2022 [citováno 3. 05. 2023]. Dostupné z: https://cs.wikipedia.org/w/index.php?title=Teorie_pravd%C4%9Bpodobnosti&oldid=21702647
- [6] NAGY, Ivan a Pavla PECHERKOVÁ. *Statistika* [online]. Praha: FD ČVUT [cit. 2023-05-02]. Dostupné z: <http://staff.utia.cas.cz/uglickich/pdfka/Statistika.pdf>
- [7] Rovnoměrné rozdělení. *Simulace.info* [online]. [cit. 2023-05-02]. Dostupné z: https://www.simulace.info/index.php/Uniform_distribution/cs
- [8] *Wikipedie: Otevřená encyklopedie: Pravděpodobnostní funkce* [online]. c2022 [citováno 3. 05. 2023]. Dostupné z: https://cs.wikipedia.org/w/index.php?title=Pravd%C4%9Bpodobnostn%C3%AD_funkce&oldid=21593478
- [9] SCUTARI, Marco a Jean-Baptiste DENIS, 2021. *Bayesian Networks* [online]. Boca Raton: Chapman and Hall/CRC [cit. 2023-05-03]. ISBN 9780429347436. Dostupné z: doi:10.1201/9780429347436
- [10] HORNÝ, Michal. Bayesian networks. Boston University School of Public Health, 2014, 17.
- [11] VANIŠ, Miroslav, 2021. Optimization of Bayesian networks and their prediction properties [online]. Praha [cit. 2023-05-03]. Dostupné z: <https://dspace.cvut.cz/bitstream/handle/10467/98538/F6-D-2021-Vanis-Miroslav.pdf?sequence=-1&isAllowed=y>. Disertační práce. ČVUT v Praze, Fakulta dopravní, Ústav aplikované matematiky.
- [12] DE OÑA, Juan, Randa Oqab MUJALLI a Francisco J. CALVO, 2011. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. *Accident Analysis & Prevention* [online]. **43**(1), 402-411 [cit. 2023-05-03]. ISSN 00014575. Dostupné z: doi:10.1016/j.aap.2010.09.010
- [13] Introduction to Bayesian Networks. *Towards Data Science* [online]. [cit. 2023-05-02]. Dostupné z: <https://towardsdatascience.com/introduction-to-bayesian-networks-81031eed94e>
- [14] Introduction to R. *R project* [online]. [cit. 2023-05-03]. Dostupné z: <https://www.r-project.org/about.html>
- [15] Okresy. *Krajská správa ČSÚ pro Středočeský kraj* [online]. [cit. 2023-05-03]. Dostupné z: <https://www.czso.cz/csu/xs/okresy>
- [16] Úvod. *Krajská správa ČSÚ v hl. m. Praze* [online]. [cit. 2023-05-03]. Dostupné z: <https://www.czso.cz/csu/xa>
- [17] *Ročenka dopravy 2021* [online], 2022. Praha: Technická správa komunikací hl. m. Prahy [cit. 2023-05-03]. Dostupné z: <https://www.tsk-praha.cz/static/udi-rocenka-2021-cz.pdf>
- [18] Charakteristika kraje. *Krajská správa ČSÚ v hl. m. Praze* [online]. [cit. 2023-05-03]. Dostupné z: https://www.czso.cz/documents/10180/202233222/33012022_charCZ.pdf/93aaa06b-d7a6-47ed-8d87-192ceb8e6aa1?version=1.1

- [19] Okres Praha-východ. *Krajská správa ČSÚ pro Středočeský kraj* [online]. [cit. 2023-05-03]. Dostupné z: https://www.czso.cz/documents/11240/17822577/okr_praha_vychod.pdf/9d79c686-a138-4573-ab2d-cdf420c2fa2c?version=1.21
- [20] Okres Praha-západ. *Krajská správa ČSÚ pro Středočeský kraj* [online]. [cit. 2023-05-03]. Dostupné z: https://www.czso.cz/documents/11240/17822577/okr_praha_zapad.pdf/5a0ba5f4-27fb-407b-a443-118c4b3c95e3?version=1.19
- [21] VANIS, Miroslav a Krzysztof URBANIEC, 2017. Employing Bayesian Networks and conditional probability functions for determining dependences in road traffic accidents data. In: *2017 Smart City Symposium Prague (SCSP)* [online]. IEEE, 2017, s. 1-5 [cit. 2023-05-03]. ISBN 978-1-5386-3825-5. Dostupné z: doi:10.1109/SCSP.2017.7973842
- [22] R - Factors. *Tutorials Point* [online]. [cit. 2023-05-03]. Dostupné z: https://www.tutorialspoint.com/r/r_factors.htm
- [23] Cut function. RDocumentation [online]. [cit. 2023-05-03]. Dostupné z: <https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/cut>
- [24] SKÁKALOVÁ, Barbora. *Analýza diskrétních nehodových dat*. Praha: ČVUT FD, 2017.
- [25] HONGGUO, Xu, Zhang HUIYONG a Zong FANG, 2010. Bayesian Network-Based Road Traffic Accident Causality Analysis. In: *2010 WASE International Conference on Information Engineering* [online]. IEEE, 2010, s. 413-417 [cit. 2023-05-03]. ISBN 978-1-4244-7506-3. Dostupné z: doi:10.1109/ICIE.2010.276
- [26] ALEMAYEHU, Tabor Feyissa, 2009. Bayesian Approach for Analysis of Road Traffic Accident (The Case of Addis Ababa). Addis Ababa. Diplomová práce. Addis Ababa University.
- [27] MUJALLI, Randa Oqab a Juan DE OÑA, 2011. A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks. *Journal of Safety Research* [online]. **42**(5), 317-326 [cit. 2023-05-03]. ISSN 00224375. Dostupné z: doi:10.1016/j.jsr.2011.06.010
- [28] STRIEGLER, Radim a Jiří AMBROS. *Multifaktorová analýza dopravní nehodovosti: metodika provádění*. [Brno]: Centrum dopravního výzkumu, c2014. ISBN 978-80-88074-01-4.

Seznam příloh

- Příloha 1 Kód pro zpracování dat, vytvoření sítě a zjištění výstupů (.R)
- Příloha 2 Výsledná Bayesovská síť (.xdsl)
- Příloha 3 Celková nejvyšší a nejvyšší absolutní změna všech stavů po zavedení evidence (.xlsx)

Seznam tabulek

tab. 1 - Příklad zaznamenaných veličin [zdroj vlastní]	14
tab. 2 - Marginální pravděpodobnost P(D) [zdroj vlastní].....	21
tab. 3 - Marginální pravděpodobnost P(A) [zdroj vlastní].....	21
tab. 4 - Podmíněná pravděpodobnost P(P D,A) [zdroj vlastní]	21
tab. 5 - Sdružená pravděpodobnost P(D,A) [zdroj vlastní].....	22
tab. 6 - Vypočtená marginální pravděpodobnost P(P) [zdroj vlastní]	22
tab. 7 - Vybrané údaje o hl. m. Praze k 31.12.2021 [17].....	28
tab. 8 - Počet nehod, zranění a relativní nehodovost v Praze, 2011-2021 [17].....	29
tab. 9 - Počty nehod v okresech Praha-východ a Praha-západ, 2011-2021 [2]	31
tab. 10 - Počet zaznamenávaných položek pro jednotlivá období [zdroj vlastní].....	32
tab. 11 - Sloučení parametrů p23 – řízení provozu v době nehody a p24 – místní úprava přednosti v jízdě [zdroj vlastní].....	34
tab. 12 - Četnosti a nové zařazení hodnot parametru Den v týdnu [zdroj vlastní].....	36
tab. 13 - Četnosti a nové zařazení hodnot parametru Čas [zdroj vlastní].....	36
tab. 14 - Četnosti a nové zařazení hodnot parametru Datum (p2a) [zdroj vlastní]	37
tab. 15 - Četnosti hodnot parametru Lokalita nehody (p5a) [zdroj vlastní]	37
tab. 16 - Četnosti a nové zařazení hodnot parametru Druh nehody (p6) [zdroj vlastní]	38
tab. 17 - Četnosti a nové zařazení hodnot parametru Druh srážky jedoucích vozidel (p7) [zdroj vlastní]	38
tab. 18 - Četnosti a nové zařazení hodnot parametru Druh pevné překážky (p8) [zdroj vlastní].....	39
tab. 19 - Četnosti a hodnot parametru Charakter nehody (p9) [zdroj vlastní]	39
tab. 20 - Četnosti a nové zařazení hodnot parametru Zavinění nehody (p10) [zdroj vlastní]	40
tab. 21 - Četnosti a nové zařazení hodnot parametru Alkohol u viníka (p11) [zdroj vlastní]	40
tab. 22 - Četnosti hodnot parametru Hlavní příčina (p12) [zdroj vlastní].....	41
tab. 23 - Četnosti a nové zařazení hodnot parametru Druh povrchu vozovky (p15) [zdroj vlastní].....	41
tab. 24 - Četnosti a nové zařazení hodnot parametru Stav povrchu vozovky (p16) [zdroj vlastní]	42
tab. 25 - Četnosti a nové zařazení hodnot parametru Stav komunikace (p17) [zdroj vlastní]	42
tab. 26 - Četnosti a nové zařazení hodnot parametru Povětrnostní podmínky (p18) [zdroj vlastní] ...	43
tab. 27 - Četnosti a nové zařazení hodnot parametru Viditelnost (p19) [zdroj vlastní]	43
tab. 28 - Četnosti a nové zařazení hodnot parametru Rozhledové poměry (p20) [zdroj vlastní].....	44
tab. 29 - Četnosti a nové zařazení hodnot parametru Dělení komunikace (p21) [zdroj vlastní]	44
tab. 30 - Četnosti a nové zařazení hodnot parametru Situování nehody na komunikaci (p22) [zdroj vlastní].....	45
tab. 31 - Četnosti a nové zařazení hodnot parametru Specifická místa a objekty (p27) [zdroj vlastní] 45	45
tab. 32 - Četnosti a nové zařazení hodnot parametru Směrové poměry (p28) [zdroj vlastní]	46
tab. 33 - Četnosti a nové zařazení hodnot parametru Místo nehody (p35) [zdroj vlastní]	46
tab. 34 - Četnosti a nové zařazení hodnot parametru Druh komunikace (p36) [zdroj vlastní]	47
tab. 35 - Četnosti a nové zařazení hodnot parametru Druh vozidla (p44) [zdroj vlastní].....	47
tab. 36 - Četnosti hodnot parametru Smyk (p49) [zdroj vlastní]	48
tab. 37 - Četnosti a nové zařazení hodnot parametru Únik hmot (p50b) [zdroj vlastní].....	48
tab. 38 - Četnosti a nové zařazení hodnot parametru Kategorie řidiče (p55a) [zdroj vlastní].....	49
tab. 39 - Četnosti hodnot parametru Nejvyšší ukončené vzdělání (p55b) [zdroj vlastní].....	49
tab. 40 - Četnosti hodnot parametru Délka řídičské praxe v řízení vozidla (p56) [zdroj vlastní].....	50
tab. 41 - Četnosti a nové zařazení hodnot parametru Stav řidiče (p57) [zdroj vlastní]	50
tab. 42 - Četnosti a nové zařazení hodnot parametru Vnější ovlivnění řidiče (p58) [zdroj vlastní].....	51
tab. 43 - Rodiče a děti uzlů vytvořené Bayesovské sítě [zdroj vlastní]	53
tab. 44 - Stav s největší kladnou změnou po zavedení evidence [zdroj vlastní]	56
tab. 45 - Stav s největší zápornou změnou po zavedení evidence [zdroj vlastní]	57
tab. 46 - Stav s největší absolutní změnou po zavedení evidence [zdroj vlastní]	58

Seznam obrázků

obr. 1 - Příklad neorientovaného grafu [zdroj vlastní]	9
obr. 2 - Příklad orientovaného grafu [zdroj vlastní]	9
obr. 3 - Příklad cyklu v grafu [zdroj vlastní]	9
obr. 4 - Příklad rodičů a dětí v grafu [zdroj vlastní]	10
obr. 5 - Příklad předchůdců a následníků v grafu [zdroj vlastní]	10
obr. 6 - Distribuční funkce diskrétní a spojitě veličiny [7]	13
obr. 7 - Příklad pravděpodobnostní funkce [8]	13
obr. 8 - Schéma příkladu podmíněné závislosti [zdroj vlastní]	17
obr. 9 - Schéma pro příklad podmíněné nezávislosti [zdroj vlastní]	18
obr. 10 - Příklad Bayesovské sítě vytvořené v softwaru GeNIe [zdroj vlastní]	20
obr. 11 - Příklad sítě s pravděpodobnostmi [zdroj vlastní]	21
obr. 12 - Síť před zavedením evidence [zdroj vlastní]	23
obr. 13 - Síť po zavedení evidence [zdroj vlastní]	23
obr. 14 - Mapa Středočeského kraje [15]	27
obr. 15 - Vytvořená Bayesovská síť [zdroj vlastní]	52

Seznam grafů

graf 1 - Počet nehod na území hl. m. Prahy, 2011-2021 [17].....	30
graf 2 - Počty zranění při nehodách na území hl. m. Prahy, 2011-2021 [17].....	30
graf 3 - Počet nehod v okresech Praha-východ a Praha-západ, 2011-2021 [2]	31