

Tato prezentace je součástí wiki-prezentace [Metoda GUHA, LISp-Miner a typové úlohy](#)

Je dostupná z [této adresy](#)

Verse 22. 8. 2019

Typ úlohy: snížení počtu dopravních nehod

Data: [Accidents](#)

Problém: *Za jakých okolností nehody způsobí změna maximální rychlosti násobné snížení pravděpodobnosti nehody pro některý stupeň závažnosti? Tyto vztahy nás zajímají obecně i pro skupiny řidičů dané věkem a pohlavím.*

Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

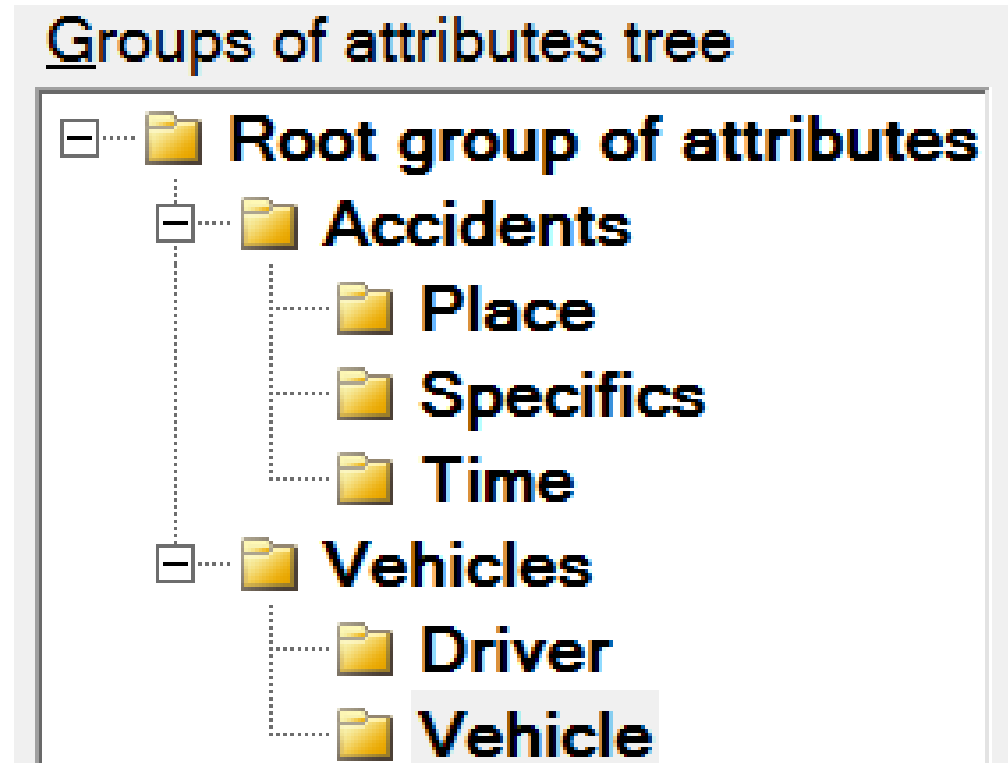
© Jan Rauch

Snížení počtu dopravních nehod - data (1)

- UK Car Accidents 2005-2015 - Data from the UK Department for Transport
<https://www.kaggle.com/silicon99/dft-accident-data>
- Data předzpracována pomocí systému LISp-Miner týmem 02 v rámci předmětu 4IZ460 v zimním semestru 2018/2019
- Původní dataset omezen pouze na nehody jednoho vozidla, celkem 538 989 řádků v analyzované matici dat
- Úlohy týkající se snížení počtu dopravních nehod řešeny týmem 27

Snížení počtu dopravních nehod - data (2)

- Vytvořeny skupiny atributů



- Cca 40 atributů celkem

Snížení počtu dopravních nehod - data (3)

Ukázky atributů

Time	Con, 0 - 1
» Day_of_Week (seq), 1 - 3	B, pos
» Month (subset), 1 - 1	B, pos
» Quarter (subset), 1 - 1	B, pos
» Rush_hours (subset), 1 - 1	B, pos
» Weekend (subset), 1 - 1	B, pos

Vehicle	Con, 0 - 1
» Age_of_Vehicle_ef (seq), 1 - 2	B, pos
» Propulsion_Code 5000+ (subset), 1 - 1	B, pos
» Vehicle_Type 5000 + (subset), 1 - 1	B, pos
» Vehicle_Type_Groups 5 000 + (subset),	B, pos

Place	Con, 0 - 1
» Area_type (subset), 1 - 1	B, pos

Driver	Con, 0 - 1
» Age_Groups_of_Driver (seq), 1 - 2	B, pos
» Driver_Home_Area_Type#2 (subset), 1 -	B, pos
» Driver_IMD_Decile (seq), 1 - 2	B, pos
» Journey_Purpose_of_Driver (subset), 1 -	B, pos
» Sex_of_Driver (subset), 1 - 1	B, pos

Specifics	Con, 0 - 1
» 1st_Road_Class (subset), 1 - 1	B, pos
» Accident_Severity (subset), 1 - 1	B, pos
» Carriageway_Hazards 5 000+ (subset), 1	B, pos
» Did_Police_Officer_Attend_Scene_of_Ac	B, pos
» Junction_Detail (subset), 1 - 1	B, pos
» Light_Conditions (subset), 1 - 1	B, pos
» Number_of_Casualties_ef5 5000 + (subs	B, pos
» Road_Surface_Conditions 5000 + (subs	B, pos
» Speed_limit_ed3 (subset), 1 - 1	B, pos
» Weather_Conditions 5 000 + (subset), 1	B, pos

Snížení počtu nehod - analytická otázka (1)

Za jakých okolností nehody způsobí změna maximální rychlosti násobné snížení pravděpodobnosti nehody pro některý stupeň závažnosti.

Tyto vztahy nás zajímají obecně i pro skupiny řidičů dané věkem a pohlavím.

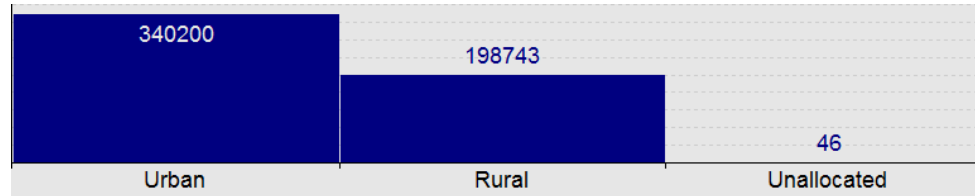
ANTECEDENT STABLE PART

Okolnosti nehody

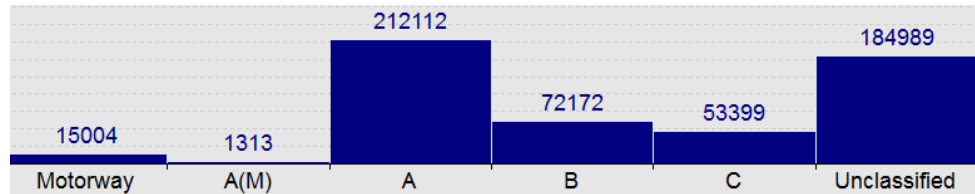
Place of accident	Con, 0 - 2
» Area_type (subset), 1 - 1	B, pos
» 1st_Road_Class (subset), 1 - 1	B, pos
Time of accident	Con, 0 - 2
» Day_of_Week (subset), 1 - 1	B, pos
» Rush_hours (subset), 1 - 1	B, pos
Car	Con, 0 - 1
» Engine_Capacity__CC_ (subset), 1 -	B, pos

Okolnosti nehody

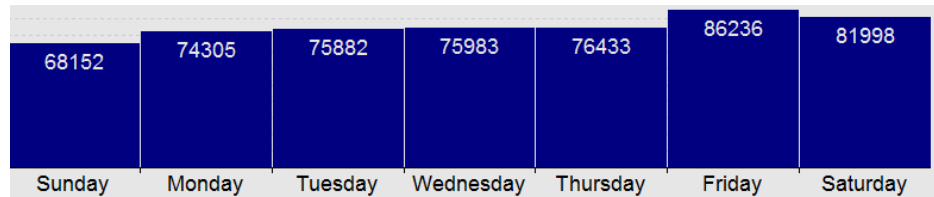
Area_type



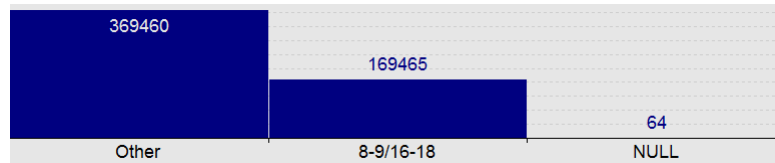
1 st Road_Class



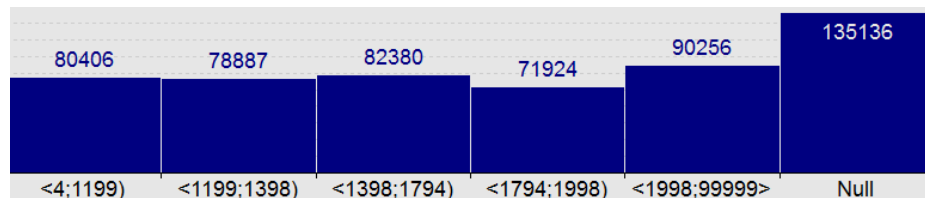
Day_of_Week



Rush_Hours



Engine_Capacity_CC



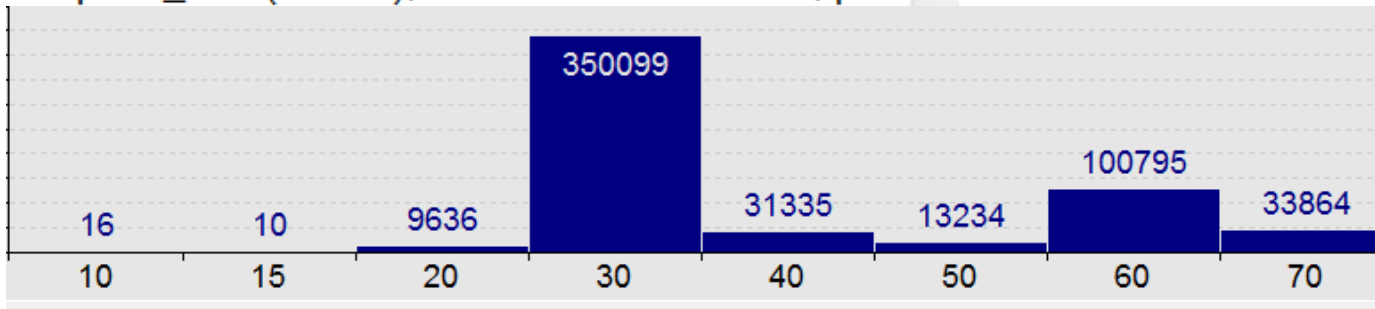
Antecedent variable part, Succedent stable part

(1) ANTECEDENT VARIABLE PART

Změna maximální rychlosti

Default Partial Cedent
» Speed_limit (subset), 1 - 1

Con, 1 - 1
B, pos

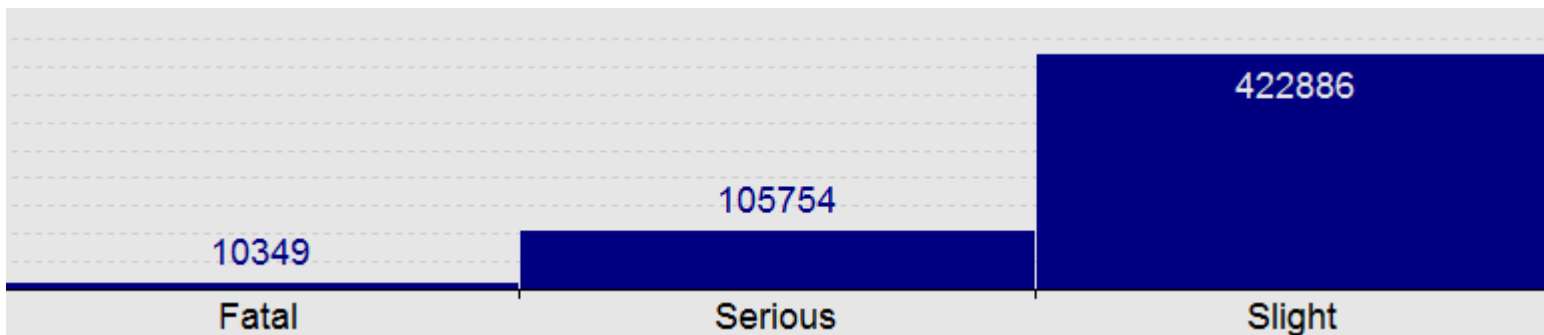


SUCCEDED STABLE PART

Závažnost nehody

Default Partial Cedent
» Accident_Severity (subset), 1 - 1

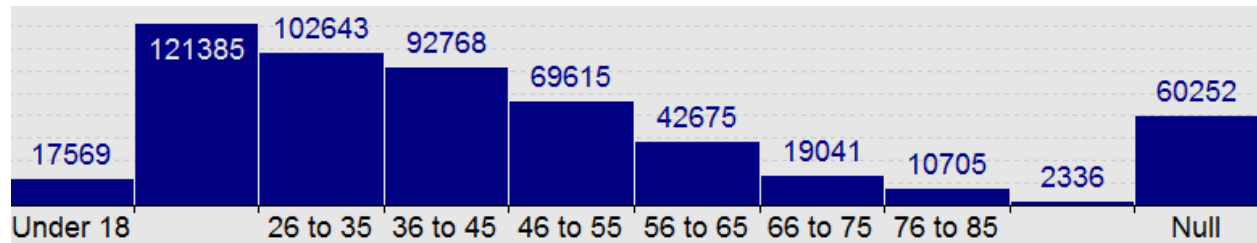
Con, 0 - 5
B, pos



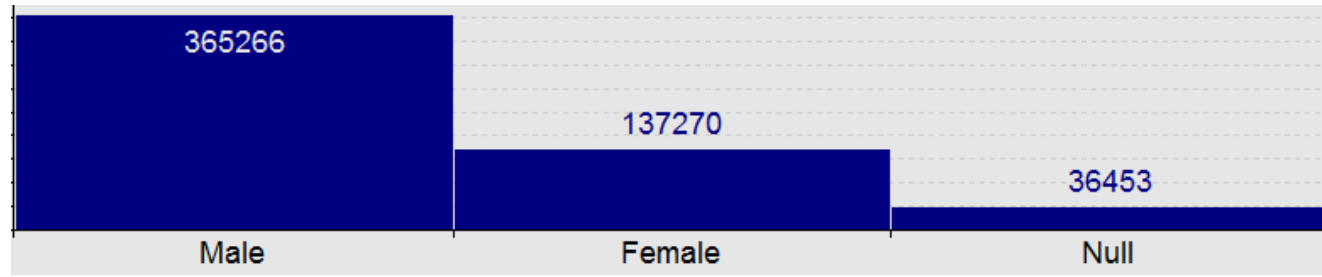
Condition

<u>CONDITION</u>	Skupiny řidičů
Default Partial Cedent	Con, 0 - 5
» Age_Groups_of_Driver (subset), 1 - 1	B, pos
» Sex_of_Driver (subset), 1 - 1	B, pos

Age_Groups



Sex



Ac4ft-kvantifikátor

Čtyřpolní tabulka před změnou rychlosti

	Závažnost	\neg Závažnost
$Okolnosti \wedge Rychlost_před \wedge Skupina$	a_I	b_I
$\neg (Okolnosti \wedge Rychlost_před \wedge Skupina)$	c_I	d_I

Čtyřpolní tabulka po změně rychlosti

	Závažnost	\neg Závažnost
$Okolnosti \wedge Rychlost_po \wedge Skupina$	a_F	b_F
$\neg (Okolnosti \wedge Rychlost_po \wedge Skupina)$	c_F	d_F

$$a_I \geq 700 \wedge a_F \geq 700 \wedge$$

$$\frac{\frac{a_I}{a_I + b_I}}{\frac{a_F}{a_F + b_F}} \geq 2$$

Pravděpodobnost, že pro nehody za **Okolnosti** nastává pro **Skupinu Závažnost** je pro **Rychlost_před** nejméně dvakrát vyšší než pro **Rychlost_po**.

Ac4ft-kvantifikátor

QUANTIFIERS			
Type	Rel.	Value	Units
a (BASE) Before	>=	700.00	Abs
a (BASE) After	>=	700.00	Abs
PIM RatioVal	>=	2.00	Abs

Ac4ft Statistical quantifier settings

Interest measure type
p-Implication
 $a/(a+b) \geq p \dots$ at least $100 \cdot p$ [%] of objects satisfying A satisfy also S

Relation
Greater than or equal

Threshold value
2

Operation mode
Ratio of interest-measures
Test applied to the ratio of interest-measures computed separately from each frequency table

Primary IM Settings
 Set as primary IM
 Normalize value range: From 0 To 1

Note
-

OK Cancel

Zadání pro Ac4ft-Miner

ANTECEDENT STABLE PART	QUANTIFIERS	SUCCEDENT STABLE PART																
Place of accident Con, 0 - 2 » Area_type (subset), 1 - 1 B, pos » 1st_Road_Class (subset), 1 - 1 B, pos Time of accident Con, 0 - 2 » Da Okolnosti nehody B, pos » Ru B, pos Car Con, 0 - 1 » Engine_Capacity__CC_ (subset), 1 - B, pos	<table border="1"> <thead> <tr> <th>Type</th> <th>Rel.</th> <th>Value</th> <th>Units</th> </tr> </thead> <tbody> <tr> <td>a (BASE) Before</td> <td>>=</td> <td>700.00</td> <td>Abs</td> </tr> <tr> <td>a (BASE) After</td> <td>>=</td> <td>700.00</td> <td>Abs</td> </tr> <tr> <td>PIM RatioVal</td> <td>>=</td> <td>2.00</td> <td>Abs</td> </tr> </tbody> </table> Generation information Status: Solved, 10 run(s) Mode: Standard	Type	Rel.	Value	Units	a (BASE) Before	>=	700.00	Abs	a (BASE) After	>=	700.00	Abs	PIM RatioVal	>=	2.00	Abs	Default Partial Cedent Con, 0 - 5 » Accident_Severity (subset), 1 - 1 B, pos Závažnost nehody
Type	Rel.	Value	Units															
a (BASE) Before	>=	700.00	Abs															
a (BASE) After	>=	700.00	Abs															
PIM RatioVal	>=	2.00	Abs															
Total length: 1 - 3		Total length: 0 - 5 {0 - 1}																
(1) ANTECEDENT VARIABLE PART	CONDITION	(2) SUCCEDENT VARIABLE PART																
Default Partial Cedent Con, 1 - 1 » Speed_limit (subset), 1 - 1 B, pos Změna maximální rychlosti	Default Partial Cedent Con, 0 - 5 » Age_Groups_of_Driver (subset), 1 - 1 B, pos » Sex_of_Driver (subset), 1 - 1 B, pos Skupiny řidičů	Default Partial Cedent Con, 0 - 5																
Total length: 1	Total length: 0 - 5	Total length: 0																

Ac4ft-Miner - výstup

Task run

Start: 23.3.2019 21:29:11 Total time: 0h 0m 50s

Number of verifications: 40142

Number of hypotheses: 11 Mode: Standard

Add group

Del group

Edit group

Actual group of hypotheses: All hypotheses

Hypotheses in group: 11 Shown hypotheses: 11 Highlighted: 0

Nr. Id R-Conf B:Conf A:Conf Hypothesis

1	3	2.725	0.039	0.014	Rush(Other) : (Speed_limit(70) -> Speed_limit(30)) >+< Accident_Severity(Fatal)
2	9	2.682	0.047	0.017	Rush(Other) : (Speed_limit(70) -> Speed_limit(30)) >+< Accident_Severity(Fatal) Sex(Male)
3	5	2.538	0.039	0.016	1st(A) : (Speed_limit(70) -> Speed_limit(30)) >+< Accident_Severity(Fatal)
4	1	2.434	0.035	0.014	Rush(Other) : (Speed_limit(40) -> Speed_limit(30)) >+< Accident_Severity(Fatal)
5	2	2.285	0.033	0.014	Rush(Other) : (Speed_limit(60) -> Speed_limit(30)) >+< Accident_Severity(Fatal)
6	4	2.266	0.035	0.016	1st(A) : (Speed_limit(60) -> Speed_limit(30)) >+< Accident_Severity(Fatal)
7	7	2.244	0.036	0.016	Area(Rural) : (Speed_limit(70) -> Speed_limit(30)) >+< Accident_Severity(Fatal)
8	8	2.233	0.039	0.017	Rush(Other) : (Speed_limit(60) -> Speed_limit(30)) >+< Accident_Severity(Fatal) Sex(Male)
9	10	2.224	0.042	0.019	1st(A) : (Speed_limit(60) -> Speed_limit(30)) >+< Accident_Severity(Fatal) Sex(Male)
10	6	2.183	0.038	0.017	1st(A) & Rush(Other) : (Speed_limit(60) -> Speed_limit(30)) >+< Accident_Severity(Fatal)
11	11	2.144	0.044	0.021	1st(A) & Rush(Other) : (Speed_limit(60) -> Speed_limit(30)) >+< Accident_Severity(Fatal) Sex(Male)

Ac4ft-Miner - detail výstupu

Antecedent: Rush(Other) : (Speed_limit(70) -> Speed_limit(30))
Succedent: Accident_Severity(Fatal)
State before: Rush(Other) && Speed_limit(70) >+< Accident_Severity(Fatal) / Sex(Male)
State after: Rush(Other) && Speed_limit(30) >+< Accident_Severity(Fatal) / Sex(Male)
Condition: Sex(Male)

Okolnosti nehody: Rush(Other)

Změna maximální rychlosti: 70 →30

Závažnost nehody: Fatal

Skupina řidičů: Sex(Male)

TEXT | DATA | STATE BEFORE | STATE AFTER | B+A | DIFF ABS | DIFF RE

	Succedent	¬Succedent	Succedent	¬Succedent
Antecedent	836	16981	2788	156556
¬Antecedent	7778	339671	5826	200096

Pravděpodobnost před:
 $836 / (836 + 16\,981) = 0.047$

Pravděpodobnost po:
 $2788 / (2788 + 156\,656) = 0.0179$

Pravděpodobnost před je 2.682 krát vyšší

Poznámky

- Výsledky zřejmě pravdivé ale nepřekvapivé
- Bude obtížné získat prakticky zajímavé výsledky
- Důležité detaily o nehodě, nemusí být dostupné
- Příklady možná užitečných detailů viz dále

Příklady možná užitečných detailů (1)

Light_Conditions	%	Frequency	Cummul. %
Daylight	66.5 %	358282	66.5 %
Darkness - lights lit	21.9 %	118168	88.4 %
Darkness - lights unlit	0.5 %	2934	88.9 %
Darkness - no lighting	9.8 %	52923	98.8 %
Darkness - lighting un...	1.2 %	6682	100.0 %

Road_Surface_Conditions	%	Frequency	Cummul. %
Dry	65.4 %	352523	65.4 %
Wet or damp	29.9 %	161228	95.3 %
Snow	0.9 %	4834	96.2 %
Frost or ice	3.4 %	18212	99.6 %
Flood over 3cm. deep	0.3 %	1434	99.9 %
Null	0.1 %	758	100.0 %

Příklady možná užitečných detailů (2)

Junction_Detail	%	Frequency	Cummul. %
Not at junction or within 20 metres	55.7 %	300076	55.7 %
Roundabout	5.0 %	26847	60.7 %
Mini-roundabout	0.7 %	3510	61.3 %
T or staggered junction	25.9 %	139772	87.2 %
Slip road	1.2 %	6256	88.4 %
Crossroads	6.1 %	32830	94.5 %
More than 4 arms (not roundabout)	1.1 %	5785	95.6 %
Private drive or entrance	1.7 %	9385	97.3 %
Other junction	2.7 %	14523	100.0 %
Null	0.0 %	5	100.0 %

Road_Surface_Conditions	%	Frequency	Cummul. %
Dry	65.4 %	352523	65.4 %
Wet or damp	29.9 %	161228	95.3 %
Snow	0.9 %	4834	96.2 %
Frost or ice	3.4 %	18212	99.6 %
Flood over 3cm. deep	0.3 %	1434	99.9 %
Null	0.1 %	758	100.0 %