Tato prezentace je součástí wiki-prezentace Metoda GUHA, LISp-Miner a typové úlohy
Je dostupná z této adresy
Verse 15. 2. 2020

Typ úlohy: Násobné snížení konfidence dodatečnou podmínkou

Data: Hotel

Problém: *Snížení relativní četnosti extrémních hodnocení dodatečnou podmínkou*

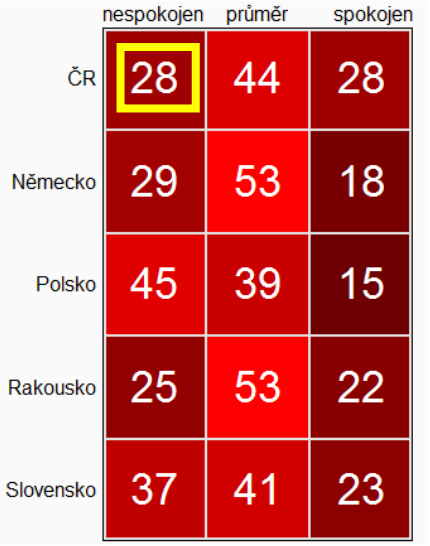
Jan Rauch
Katedra informačního a znalostního inženýrství
Vysoká škola ekonomická v Praze

Snížení relativní četnosti extrémních hodnocení dodatečnou podmínkou

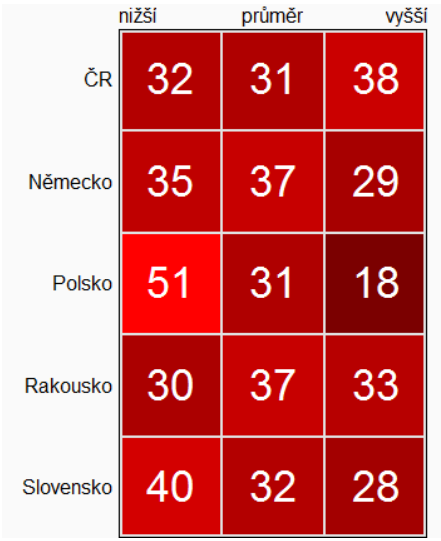
- Motivace
- Princip
- SD4ft-Miner - příklad zadání parametrů
- SD4ft-kvantifikátor
- Second set = dodatečná podmínka
- Přehled výsledků
- Nejsilnější vztah - detail
- Nejsilnější vztah - detail, poznámky

Motivace

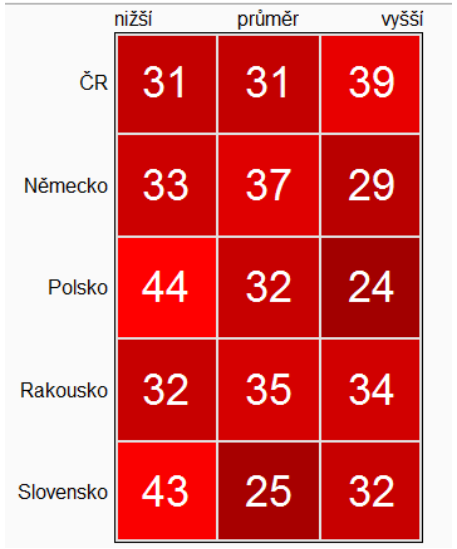
DHodnoceni



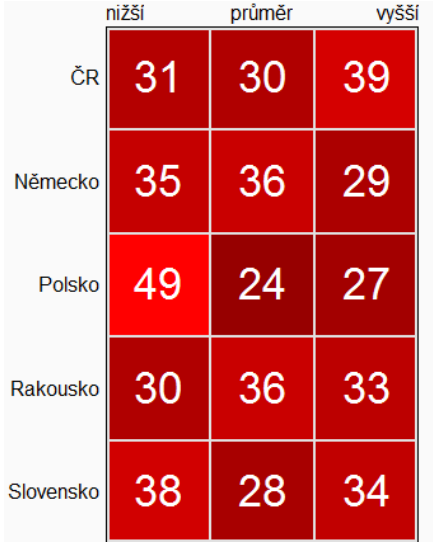
DPersonal



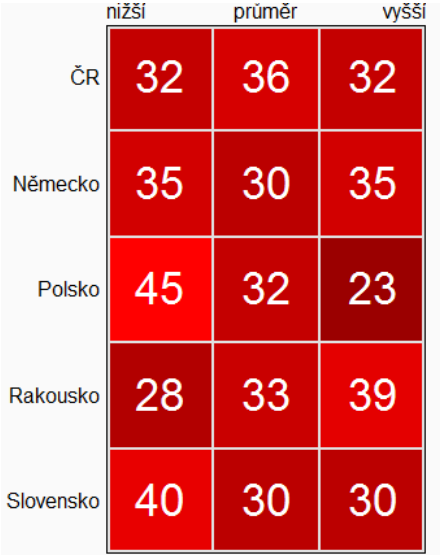
DStrava



DUbytování



DZabava



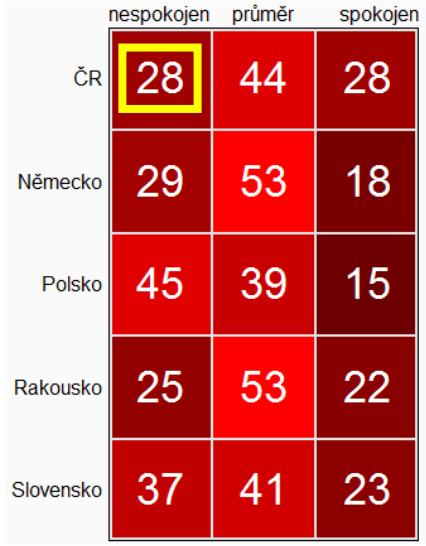
Relativní četnost nespokojených hostů z ČR je 28%. Otázka je, jaké dodatečné podmínky sníží tuto relativní četnost nespokojených hostů alespoň o jednu čtvrtinu.

Analogické otázky:
Jaké dodatečné podmínky sníží relativní četnost extrémních hodnocení u jednotlivých států alespoň o jednu čtvrtinu.

Extrémní hodnocení jsou v krajních sloupcích tabulek, jsou různá od průměrných hodnocení

Princip (1)

DHodnoceni



Relativní četnost nespokojených hostů z ČR je 28%. Otázka je, jaké dodatečné podmínky sníží tuto relativní četnost nespokojených hostů alespoň o jednu čtvrtinu.

Hotel	DHodnocení(nespokojen)	¬DHodnocení(nespokojen)
HStat(ČR)	a_1	b_1
¬HStat(ČR)	c_1	d_1

Hotel	DHodnocení(nespokojen)	¬DHodnocení(nespokojen)
HStat(ČR) ∧ Podmínka	a_2	b_2
¬(HStat(ČR) ∧ Podmínka)	c_2	d_2

Chceme: $\frac{a_2}{a_2+b_2} \leq (1-\frac{1}{4})\frac{a_1}{a_1+b_1}$, čili $\frac{\frac{a_1}{a_1+b_1}}{\frac{a_2}{a_2+b_2}} \geq 1.33$

Princip (2)

Místo pravidla HStat(ČR) ≈ DHodnocení(nespokojen) se čtyřpolní tabulkou

Hotel	DHodnocení(nespokojen)	¬DHodnocení(nespokojen)
HStat(ČR)	a_1	b_1
¬HStat(ČR)	c_1	d_1

pracujeme s podmíněným pravidlem *True* ≈ DHodnocení(nespokojen) / HStat(ČR) se čtyřpolní tabulkou

Hotel / HStat(ČR)	DHodnocení(nespokojen)	¬DHodnocení(nespokojen)
True	a_1	b_1
¬*True*	*0*	*0*

.

Zde *True* je identicky pravdivý booleovský atribut.

Princip (3)

Místo pravidla HStat(ČR) $\wedge$ Podmínka $\approx$ DHodnocení(nespokojen) se čtyřpolní tabulkou

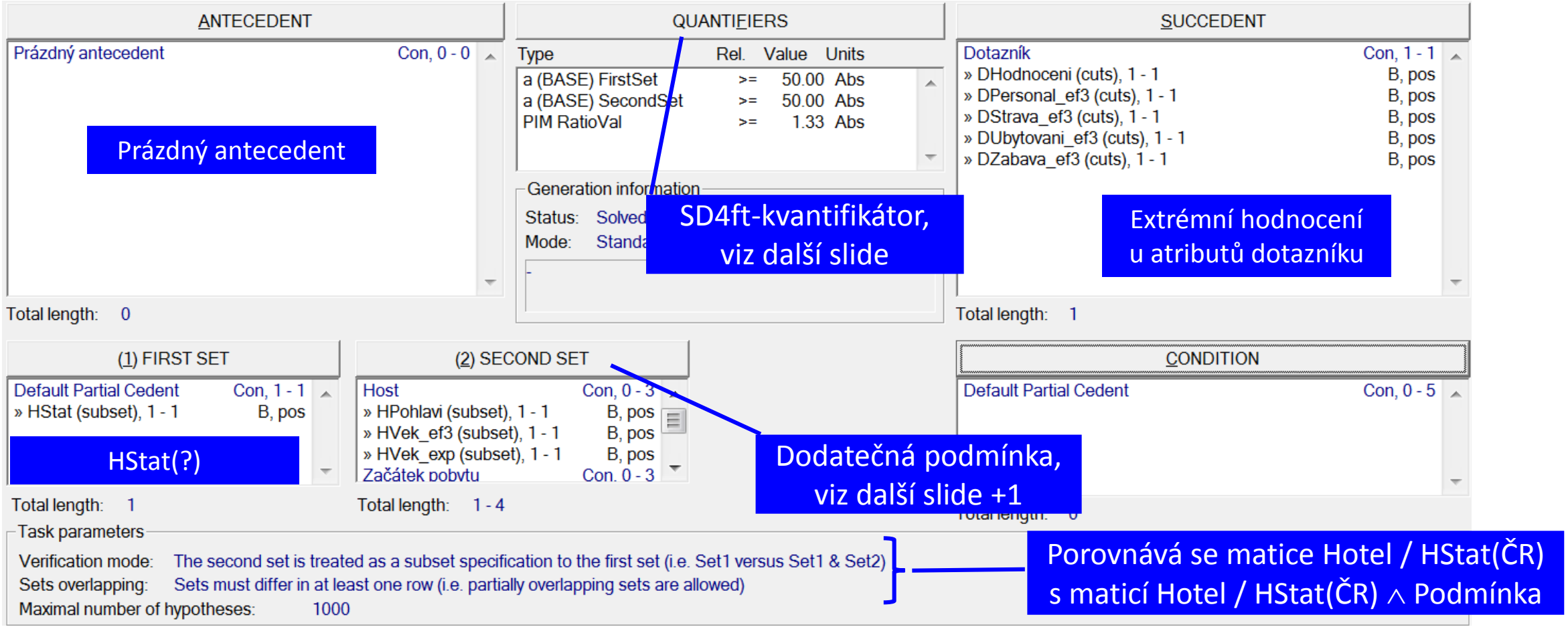
Hotel	DHodnocení(nespokojen)	¬DHodnocení(nespokojen)
HStat(ČR) $\wedge$ Podmínka	a_2	b_2
¬(HStat(ČR) $\wedge$ Podmínka)	c_2	d_2

pracujeme s podmíněným pravidlem *True* $\approx$ DHodnocení(nespokojen) / HStat(ČR) $\wedge$ Podmínka se čtyřpolní tabulkou

Hotel / HStat(ČR) $\wedge$ Podmínka	DHodnocení(nespokojen)	¬DHodnocení(nespokojen)
True	a_2	b_2
¬*True*	*0*	*0*

Zde *True* je identicky pravdivý booleovský atribut.

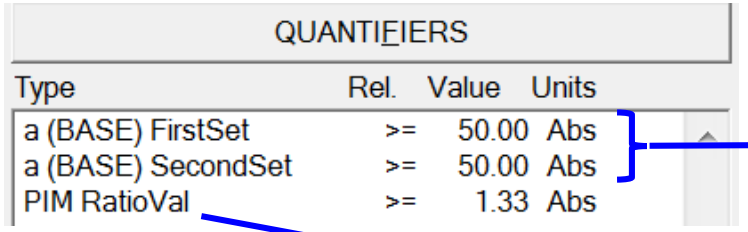
SD4ft-Miner - příklad zadání parametrů



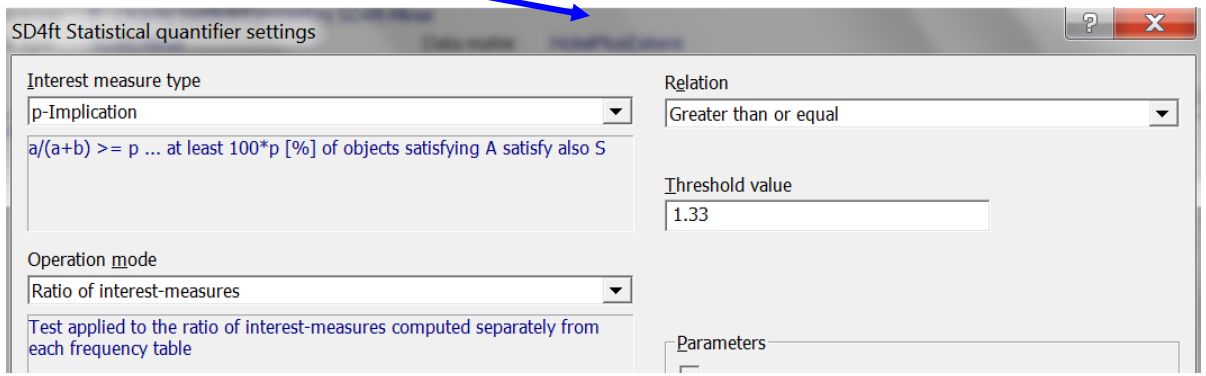
SD4ft-kvantifikátor

Hotel / HStat(ČR)	DHodnocení(nespokojen)	¬DHodnocení(nespokojen)
True	a_1	b_1
¬True	*0*	*0*

Hotel / HStat(ČR) ∧ Podmínka	DHodnocení(nespokojen)	¬DHodnocení(nespokojen)
True	a_2	b_2
¬True	*0*	*0*

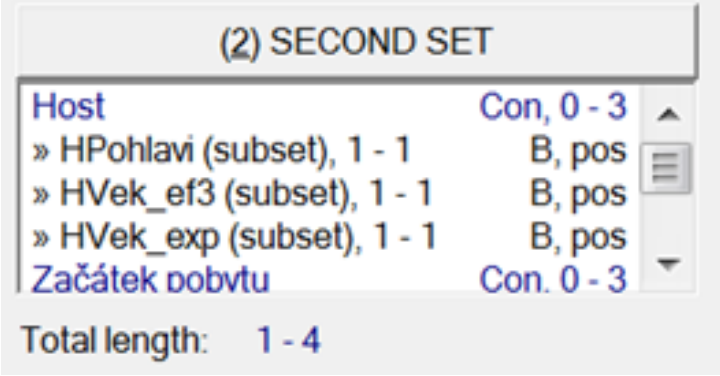


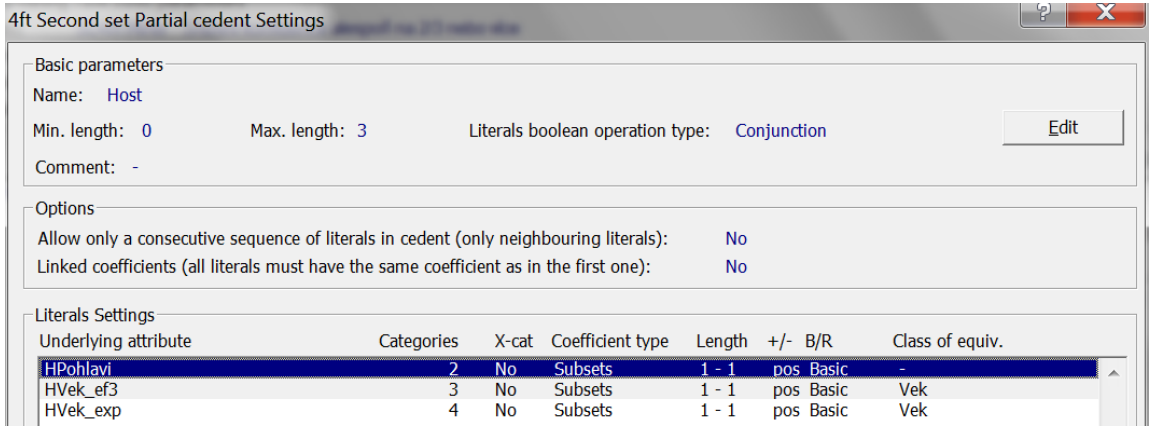
$a_1 \geq 50 \wedge a_2 \geq 50$

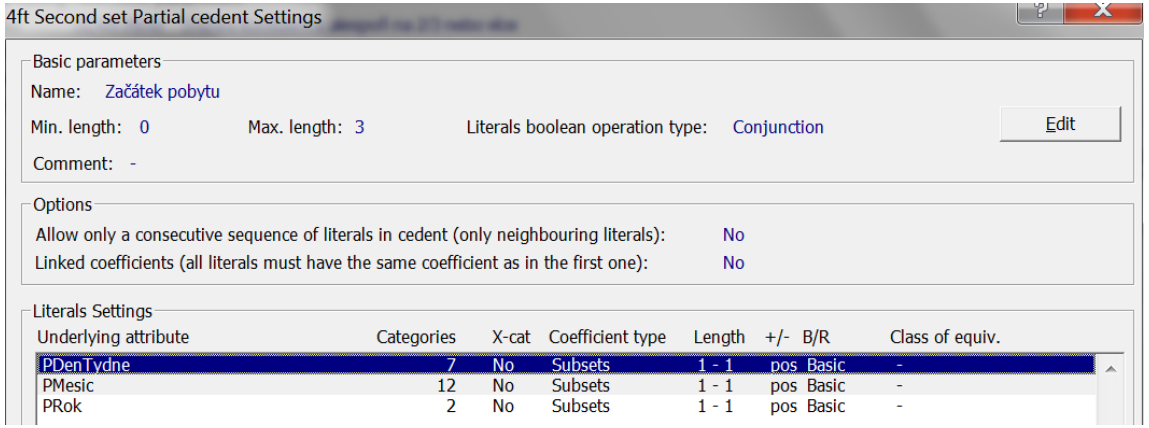


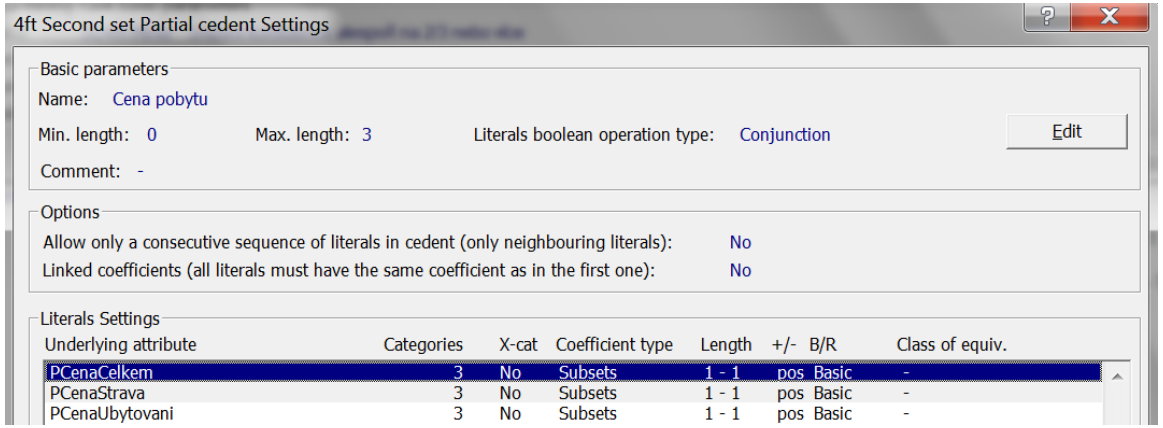
Chceme: $\frac{a_2}{a_2+b_2} \leq (1-\frac{1}{4})\frac{a_1}{a_1+b_1}$, čili $\frac{\frac{a_1}{a_1+b_1}}{\frac{a_2}{a_2+b_2}} \geq 1.33$

Second set = dodatečná podmínka









Přehled výsledků

Task run

Start: 16.2.2020 11:22:23 Total time: 0h 0m 5s

Number of verifications: 198850

Number of hypotheses: 12 Mode: Standard

Add group | Del group | Edit group

Actual group of hypotheses: All hypotheses

Hypotheses in group: 12 Shown hypotheses: 12 Highlighted: 0

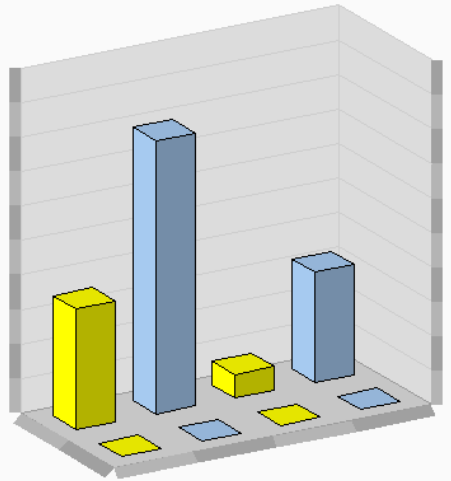
Nr.	Id	R-Conf	1:Conf	2:Conf	Hypothesis
1	6	1.722	0.306	0.177	(empty) >÷< **DStrava**(*nižší*) : **HStat**(*ČR*) × Set1 & **PRok**(*2012*) & **PCenaStrava**(*low*)
2	4	1.655	0.306	0.185	(empty) >÷< **DStrava**(*nižší*) : **HStat**(*ČR*) × Set1 & **PRok**(*2012*)
3	3	1.551	0.314	0.203	(empty) >÷< **DUbytovani**(*nižší*) : **HStat**(*ČR*) × Set1 & **PRok**(*2013*) & **PCenaStrava**(*low*)
4	5	1.531	0.324	0.212	(empty) >÷< **DZabava**(*vyšší*) : **HStat**(*ČR*) × Set1 & **PRok**(*2012*)
5	11	1.504	0.303	0.201	(empty) >÷< **DPersonal**(*nižší*) : **HStat**(*Rakousko*) × Set1 & **PRok**(*2013*)
6	1	1.474	0.314	0.213	(empty) >÷< **DUbytovani**(*nižší*) : **HStat**(*ČR*) × Set1 & **PRok**(*2013*)
7	12	1.461	0.388	0.266	(empty) >÷< **DZabava**(*vyšší*) : **HStat**(*Rakousko*) × Set1 & **PRok**(*2012*)
8	7	1.436	0.324	0.226	(empty) >÷< **DZabava**(*vyšší*) : **HStat**(*ČR*) × Set1 & **PRok**(*2012*) & **PCenaStrava**(*low*)
9	10	1.402	0.315	0.225	(empty) >÷< **DPersonal**(*nižší*) : **HStat**(*ČR*) × Set1 & **HPohlavi**(*žena*) & **PRok**(*2013*)
10	9	1.384	0.314	0.227	(empty) >÷< **DUbytovani**(*nižší*) : **HStat**(*ČR*) × Set1 & **HPohlavi**(*muž*) & **PRok**(*2013*)
11	2	1.376	0.315	0.229	(empty) >÷< **DPersonal**(*nižší*) : **HStat**(*ČR*) × Set1 & **PRok**(*2013*) & **PCenaStrava**(*low*)
12	8	1.332	0.306	0.229	(empty) >÷< **DStrava**(*nižší*) : **HStat**(*ČR*) × Set1 & **PDenTydne**(*Pá*) & **PRok**(*2012*)

Nejsilnější vztah - detail

Antecedent: (empty)
Succedent: DStrava(nižší)
First set: HStat(ČR)
Second set: HStat(ČR) Set1+ PRok(2012) & PCenaStrava(low)
Condition: (empty)

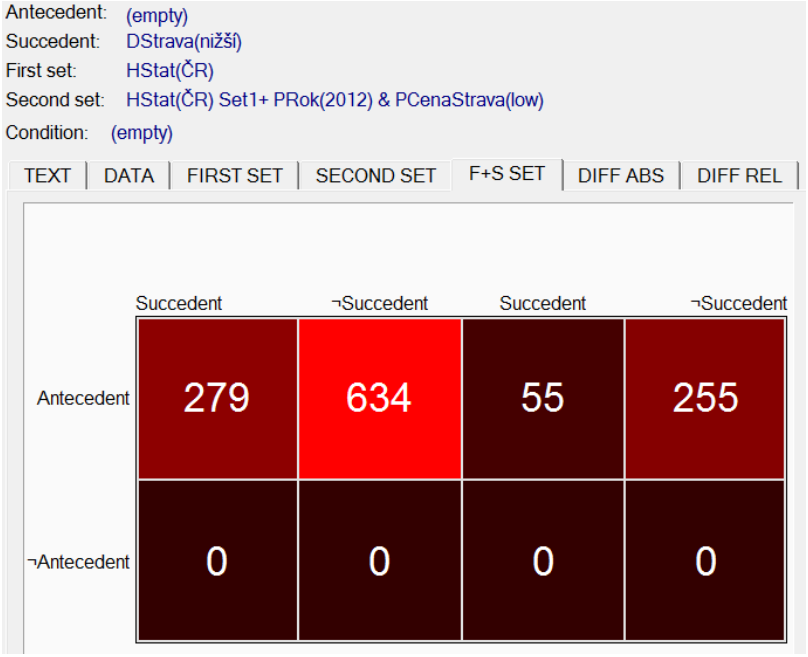
TEXT | DATA | FIRST SET | SECOND SET | F+S SET | DIFF ABS | DIFF REL

	Succedent	¬Succedent	Succedent	¬Succedent
Antecedent	279	634	55	255
¬Antecedent	0	0	0	0



- relativní četnost hodnocení DStrava(nižší) od hostů splňujících HStat(ČR) je $\frac{279}{279+634} = 0.31$
- relativní četnost hodnocení DStrava(nižší) od hostů splňujících HStat(ČR) $\wedge$ PRok(2012) $\wedge$ PCenaStrava(low) je , $\frac{55}{55+255} = 0.18$, tedy o 42% nižší.

Nejsilnější vztah - detail, poznámky



Relativní četnost hodnocení DStrava(nižší) od hostů splňujících HStat(ČR) =

- konfidence podmíněného pravidla *True* ≈ DStrava(nižší)/HStat(ČR)
- konfidence pravidla HStat(ČR) ≈ DStrava(nižší)
- charakteristika PIM podmíněného pravidla *True* ≈ DStrava(nižší)/HStat(ČR)
- charakteristika PIM pravidla HStat(ČR) ≈ DStrava(nižší)
- $\frac{279}{279+634} = 0.31.$

Relativní četnost hodnocení DStrava(nižší) od hostů splňujících od hostů splňujících HStat(ČR) ∧ PRok(2012) ∧ PCenaStrava(low) =

- konfidence podmíněného pravidla
  True ≈ DStrava(nižší)/HStat(ČR) ∧ PRok(2012) ∧ PCenaStrava(low)
- konfidence pravidla
  HStat(ČR) ∧ PRok(2012) ∧ PCenaStrava(low) ≈ DStrava(nižší)
- charakteristika PIM podmíněného pravidla
  True ≈ DStrava(nižší)/HStat(ČR) ∧ PRok(2012) ∧ PCenaStrava(low)
- charakteristika PIM pravidla
  Stat(ČR) ∧ PRok(2012) ∧ PCenaStrava(low) ≈ DStrava(nižší)
- $\frac{55}{55+255} = 0.18.$