

Dolování dat z dotazník

Ondřej Takács



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Úvod

- Součástí projektu, který se zabývá individualizovaným e-learningem
 - . virtuální učitel, který svůj výklad přizpůsobuje statickým i dynamicky se měnícím vlastnostem studenta
 - . má k dispozici:
 - bohatou zásobu různých stylů i forem výuky
 - dostatek informací o stylech učení, druhu paměti, typech inteligence a o dalších vlastnostech studenta



Styly učení

- Styl učení je souhrn stabilních vlastností ovlivňujících učení
- Analýzou publikovaných stylů učení byly vlastnosti rozděleny do následujících skupin:
 - **smyslové vnímání:** vizuální . auditivní . kinestetické - verbální
 - **sociální aspekty:** rád pracuje sám . ve dvojici . s učitелеm . ve skupině
 - **afektivní aspekty:** motivace ke studiu vnitřní, vnější
 - **taktiky učení:**
 - **systematická** s plány . volnost
 - **způsob** s plány teoretické odvozování . experimentování
 - **postup** s plány detailistický, zdola nahoru . holistický, shora dolů
 - **pojetí** s plány hloubkový . povrchový
 - **autoregulace** s plány dle pokynů - samostatně

ím se liší dotazníky od test

- Test měří úroveň znalostí a dovedností studenta
- Dotazník měří vybrané vlastnosti studenta
 - . Nejsou zde správné a špatné odpovědi
 - . Často více vlastností, nejen jedna výsledná známka
- Nový typ otázky: škála
 - . Rád se uším:
 - zcela souhlasím . souhlasím . nesouhlasím - zcela nesouhlasím

Shromáždění experimentálních dat

- Byla sestavena skupina dotazníků z různých publikovaných a ověřených dotazníků stylů
- Většina dotazníků byla zkrácena o otázky, které se netýkaly nází skupiny vlastností
- Většinu dotazníků bylo třeba přeložit
- Dotazníky měly různé způsoby vyplnění

Název dotazníku	Zjiřované vlastnosti	Nutnost p ekladu	Zkrácen	Druh otázek
LSI	Sociální aspekty, motivace k u ení, autoregulace u ení	ne	ano	řkála 1 a0 5
VARK	Smyslové vnímání	áste n	ne	Otázky s 4 textovými variantami
ILS	Systemati nost u ení (ád/volnost), zp sob u ení (teoretické odvozování/ experimentování)	ano	ano	řkála 1 a0 5
TSI	Postup u ení (analystický/holistický)	ano	ano	řkála 1 a0 7
ASSIST	Pojetí u ení (hloubkové, strategické, povrchové)	ano	ne	řkála 1 a0 5

Vyplnění dotazník

- Dotazníky byly převedeny do elektronické formy, kde byly dotazníky prezentovány samostatně
- Všechny dotazníky vyplnilo přes 500 studentů odborných a vysokozkolských studentů různých oborů
- Pro analýzu byly kromě výsledků dotazníků evidovány i odpovědi na jednotlivé otázky



Reliabilita dotazník

- Udává spolehlivost dotazníku, míru v které opakovné vyplnění dotazníku vede ke stejným výsledkům
- Při nízké reliabilitě
 - Náhodné odpovědi
 - Při opakovaném vyplnění jiné výsledky

Ur ení reliability: Cronbach alfa

- Vypo teno z vypln ných dotazník
- Porovnává variabilitu jednotlivých odpov dí s variabilitou výsledné vlastnosti

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum V_i}{V_{test}} \right)$$

- n : počet otázek m ících danou vlastnost
- V_i : variabilita odpov dí na otázky m ících danou vlastnost
- V_{test} : varibilita výsledné vlastnosti

Otázky

	I.1	I.2	I.3	I.4	I.5	I.6	I.7	I.8	I.9	I.10	M ěn ěná vlastnost	
Osoby												
S.1	0	1	0	1	1	1	0	0	1	0	5	Statistiky vlastností osob
S.2	1	0	0	0	0	0	0	1	1	0	3	Pr ům ěrn ě hodnota 6,0
S.3	1	0	0	1	1	1	0	1	1	1	7	Odchylka 2,2
S.4	1	0	1	0	1	1	0	1	0	1	6	Variabilita 5,0
S.5	1	1	1	1	1	1	0	1	1	1	9	

Variability

ot ězek 0,20 0,300,300,300,200,200,000,200,200,30 = (e.g.) VAR(L6:L10)

Suma variability = =SUM(C12:L12)

ot ězek = 2,2

=COUNT(C12:L12)

Po ět ot ězek = 10



Cronbach's alpha =

0,62

=G14/(G14-1)*(1-(G13/(O8^2))) = n/(n-1)* (1- (Suma variabilit ot ězek / (variabilita vlastností osob)))

Určení reliability: Cronbach alfa

- Závisí na tom, jak byly dotazníky vyplněny
- Hodnoty se pohybují od 0 do 1
- 0,7 a výše je přijatelná hodnota
- U malého počtu vyplněných dotazníků může i jedna odpověď velmi ovlivnit výslednou reliability

P edzpracování dat

- Uvedení dat do formy vhodné pro dolování dat
- Pokud jsou data získávána z databáze, často je nutné propojit údaje z n kolika tabulek v nové tabulce

P edzpracování: p vodní tabulka

id_vys_uziv	id_vypln_dotaz	id_vlast	hodnoceni	detaily
1	1	5	2.6	NULL
2	1	7	3.5	NULL
3	1	8	2.5	NULL
4	1	9	3.6	NULL
5	2	30	0.54545454545455	NULL
6	2	31	0.27272727272727	NULL
7	3	26	1	NULL
8	3	25	1	NULL
9	3	23	1	NULL
10	3	24	1	NULL
11	3	53	1	26,25,23,24
12	4	5	1.4	NULL

Podzpracování: vytvořená tabulka

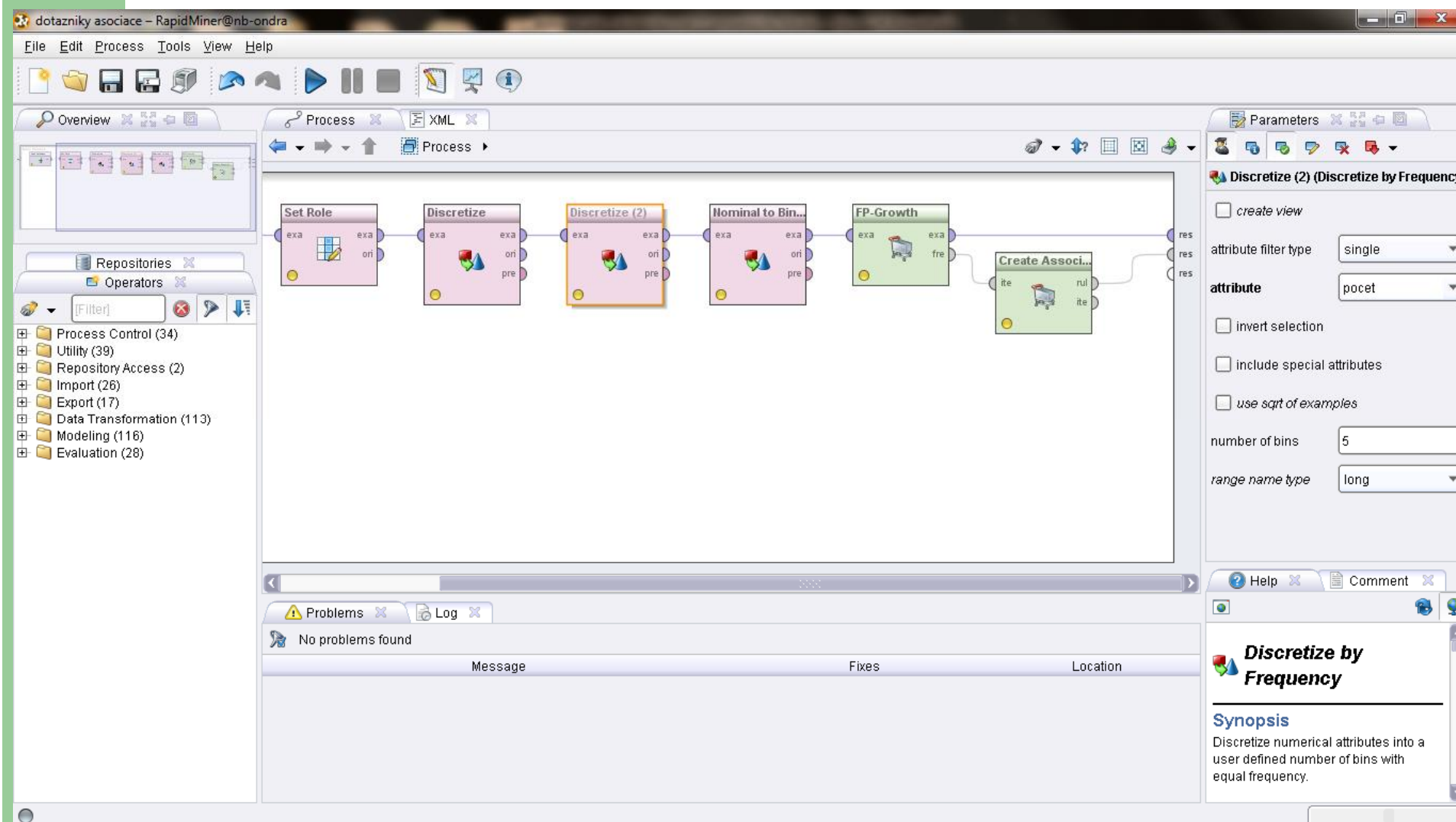
povrchový	strategický	hloubkový	autoregulace	auditivní	verbální	kinestetický	vizuální
20	35	30	45	40	20	40	0
15	10	10	35	20	20	40	20
40	35	10	40	60	0	20	20
25	20	10	35	20	20	40	20
15	20	10	45	20	20	40	20
40	10	20	55	40	60	0	0
10	5	0	55	20	20	20	40
40	60	10	45	40	60	0	0
5	0	0	25	80	20	0	0

Nástroje pro analýzy dat

- R
 - . Zdarma
 - . Ovládání p íkazovou ádkou
 - . Bohaté možnosti analýzy
- Rapidminer
 - . Zdarma
 - . Grafické rozhraní
 - . Velký výběr metod
 - . U n kterých nejdou podrobn nastavit parametry



Rapidminer



The screenshot displays the Rapidminer software interface. The main workspace shows a workflow with the following steps: Set Role, Discretize, Discretize (2), Nominal to Bin..., FP-Growth, and Create Associ... The 'Discretize (2)' step is currently selected, and its parameters are visible in the right-hand panel. The parameters for 'Discretize (2) (Discretize by Frequency)' are:

- create view
- attribute filter type: single
- attribute: pocet
- invert selection
- include special attributes
- use sqrt of examples
- number of bins: 5
- range name type: long

Below the parameter panel, there is a 'Help' section for 'Discretize by Frequency' with a synopsis: 'Discretize numerical attributes into a user defined number of bins with equal frequency.'

The interface also includes a menu bar (File, Edit, Process, Tools, View, Help), a toolbar, an Overview panel, a Repositories panel, a Problems/Log panel at the bottom, and a status bar.

Metody analýzy dat

- Pro shromážděná data byly navrženy metody jejich analýzy:
 - . Metoda hlavních komponent
 - Vytváří novou reprezentaci dat pomocí vzájemně nezávislých komponent
 - . Metoda rozhodovacích stromů
 - Hledá zajímavá pravidla, která reprezentuje formou stromu
 - . Metoda shlukování
 - Hledá skupiny podobných objektů, které se liší od ostatních skupin

Asociativní pravidla

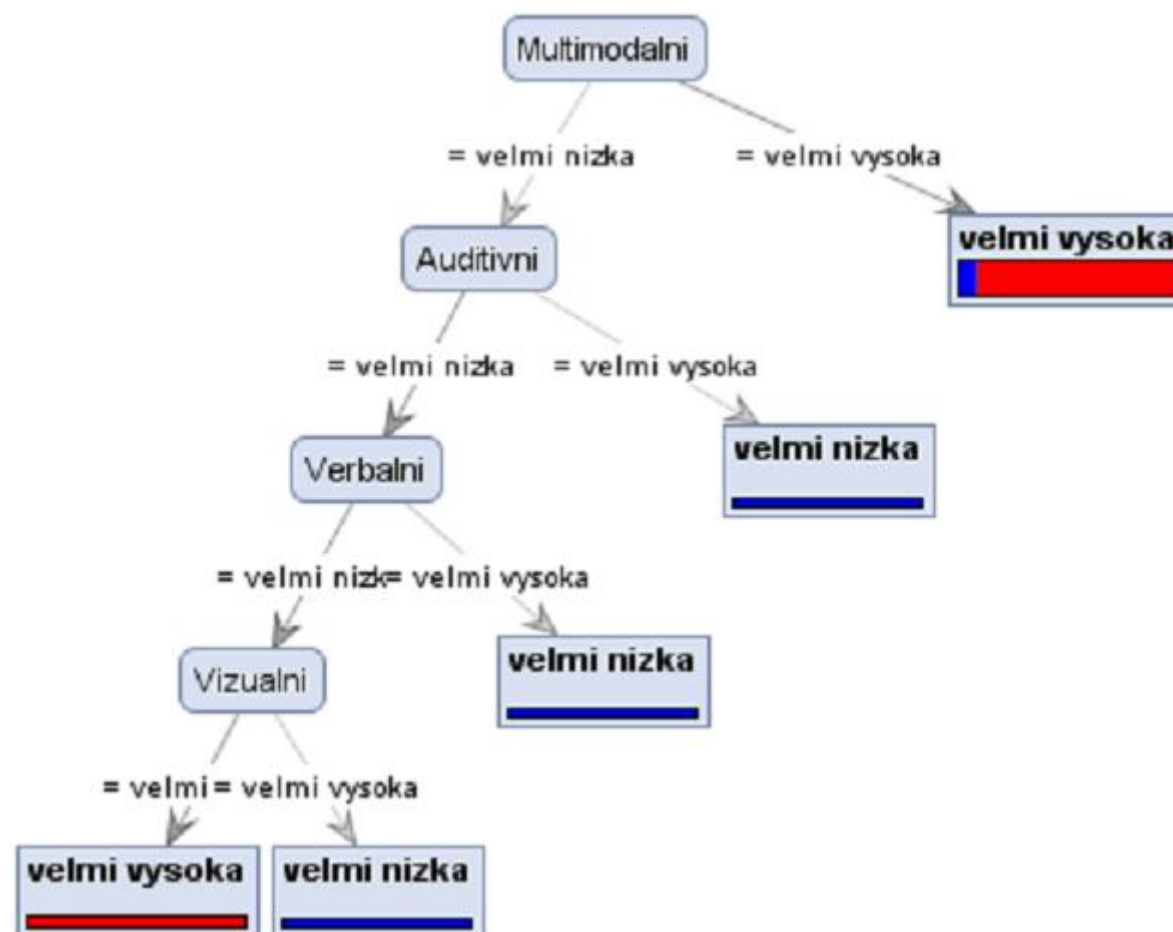
- Určuje závislosti hodnot jednotlivých atributů
- Forma pravidel:
 - Jestliže α pak β
 - Spolehlivost γ v kolika procentech p ípad platí
 - Podpora δ v kolika procentech p ípad bylo použito

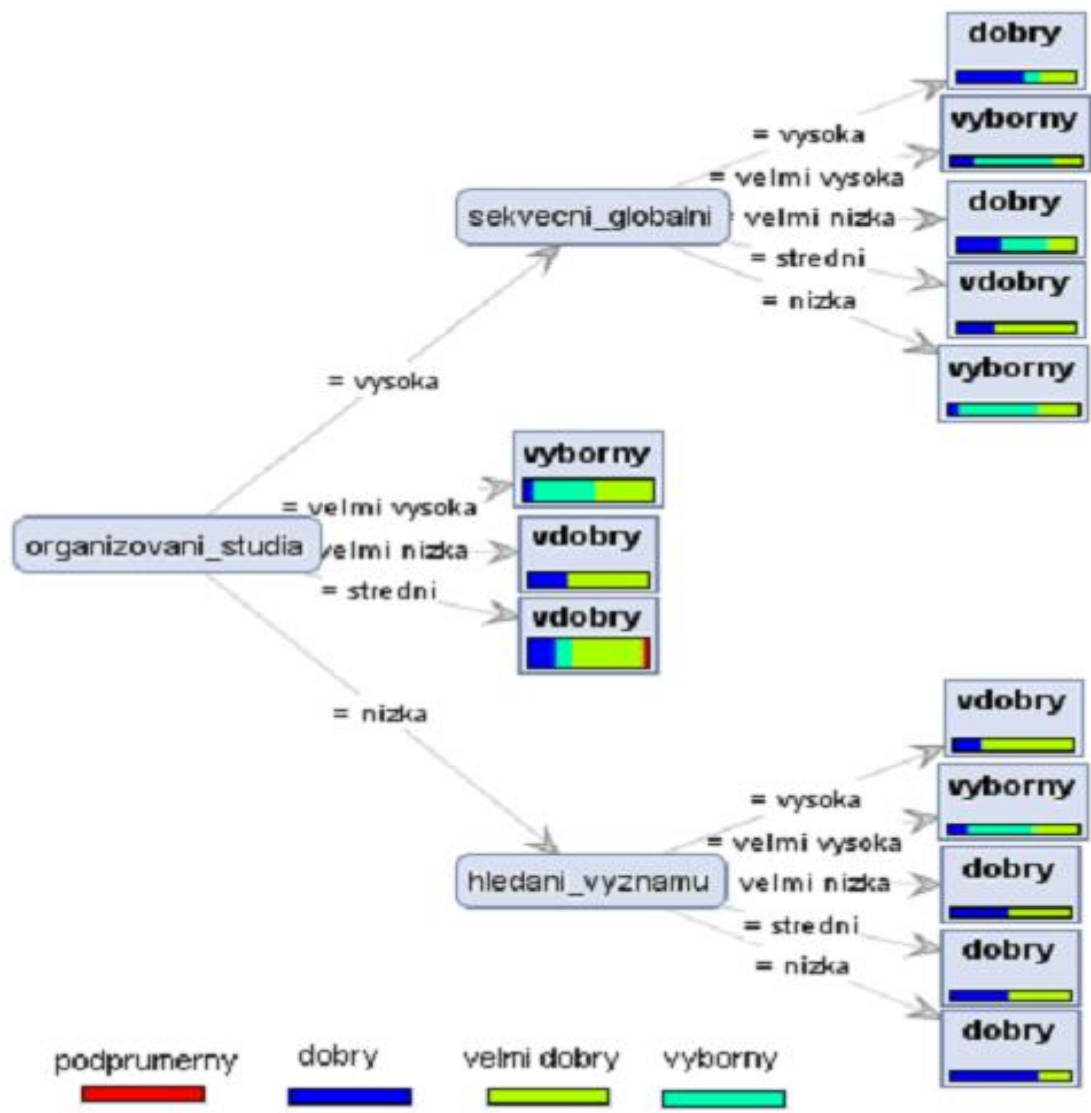


Num	Condition	Conclusion	Support	Reliability
1	visual = very low	depth = very low	43%	71%
2	strategic = very low	depth = very low	38%	76%
3	social (alone - group) = middle	depth = very low	38%	72%
4	verbal = low	depth = very low	31%	73%
5	theoretical ó experimental = high	depth = very low	31%	76%
í	í	í	í	í
32	motivation = high, theoretical ó experimental = high	depth = very low	17%	73%
33	social (alone - group) = middle, verbal = low	depth = very low	17%	75%
34	auditive = middle	visual = very low	16%	80%
í	í	í	í	í
69	depth = very low, systematic (sequential - random) = very low	strategic = very low	13%	72%
70	visual = very low, motivation = high, strategic = very low	depth = very low	13%	73%
í	í	í	í	í
87	depth = very low, visual = very low, field = University, pedagogical	motivation = high	12%	70%
88	strategic = very low, visual = middle	depth = very low	12%	79%
í	í	í	í	í
107	depth = very low, surface = very low, autoregulation = low	visual = very low	11%	71%
í	í	í	í	í
147	autoregulation = low, social (alone - group) = low	depth = very low	10%	



Analýza rozhodovacími stromy





Výsledky analýzy stylu učení

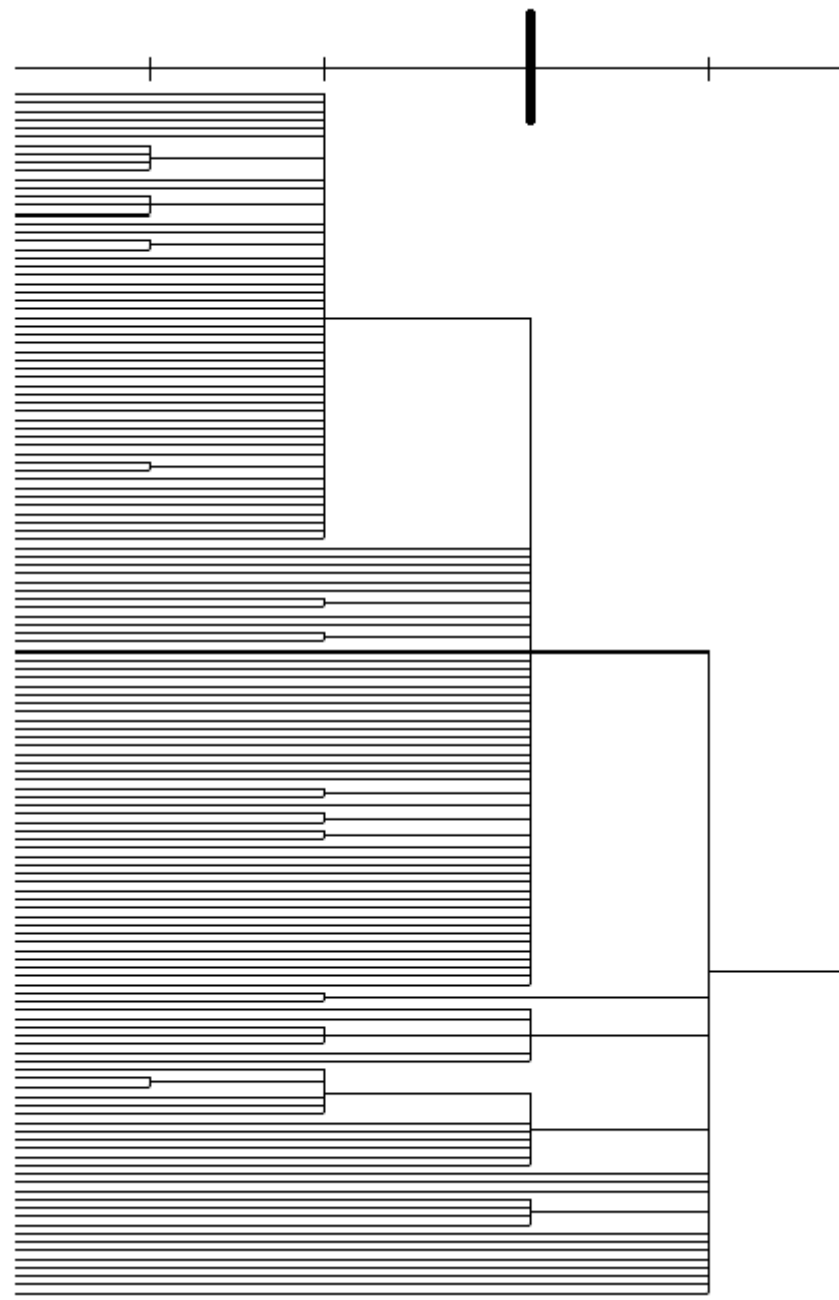
- Analýza rozhodovacími stromy
 - . Kinestetický smyslový typ je téměř úplně závislý na ostatních smyslových typech
 - . Největší vliv na kladné hodnocení studenta má dobrá organizace studia
 - . Méně zodpovědní studenti se raději učí s kamarády kdežto zodpovědní studenti se raději učí sami

Výsledky analýzy dat

- Analýza hlavních komponent
 - Počet hlavních komponent odpovídá na otázky je mnohem větší než počet hlavních komponent výsledků dotazníku. nespolehlivost shromážděných dat
 - Zvolené vlastnosti jsou na sobě závislé. bylo nalezeno 18 hlavních komponent, které mohou vyjádřit všech 28 zjižovaných vlastností, ale obtížně by se interpretovaly

Výsledky analýzy

- Analýza shlukováním
 - Nepotvrdilo se, že by existovalo n kolik skupin studentů s podobnými vlastnostmi
 - Většina studentů se svými vlastnostmi velmi liší od ostatních





D kuji za pozornost

