

## POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

**Autor/ka:** Bc. Martina Doležalová

**Název práce:** Segmentace zákazníků pro nastavení diferencovaného přístupu k nim

**Hodnocení práce:**

(hodnocení: 1 je výborný, ..., 4 je nevyhovující)

	1	1-	2	2-	3	4
1. <b>Náročnost tématu</b>			X			
2. <b>Volba vhodné metodiky zpracování</b>			X			
3. <b>Splnění cílů práce</b>			X			
4. <b>Odborný přínos (podíl) autora/rky</b>			X			

5. Logický postup řešení			X			
6. Využití teoretických znalostí			X			
7. Návrh modelu řešení			X			
8. Sběr a předzpracování dat pro řešení			X			
9. Výstižný souhrn práce			X			
10. Průběžná citace použité literatury		X				
11. Formální úprava textu				X		
12. Formální úprava grafická (obrázky, ...)		X				
13. <b>Celkové hodnocení práce</b>			X			

**Další připomínky, vyiáždění, doporučení nebo nedoporučení k obhajobě:**

Cílem práce je segmentace zákazníků vybranými data miningovými metodami. Segmentace je realizována na zákaznících IT firmy pomocí standardních metod implementovaných v programovém prostředí SPSS. Předzpracování dat pro segmentaci je provedeno pečlivě, proto lze také konstatovat, že výsledky shlukování jsou spolehlivé a vzhledem k připravené implementaci algoritmu do informačního systému by mohly být pro firmu přínosné.

K práci mám následující připomínky:

- Úvod práce by mohl být zpracován pečlivěji. Není zde odůvodněna použitá metodika ani návaznost jednotlivých kapitol. Provázanost kapitol není pak zřejmá ani ze samotného textu.
- V kapitole 1 postrádám původní citaci na metodiku CRISP-DM.
- V kapitole 2 chybí matematické zápisy všech algoritmů, včetně algoritmu učení Kohonenovy samoorganizující se mapy. Ani z dalšího textu není jasné, který algoritmus byl pro toto učení použit, a jak byly nastaveny parametry učení. Z textu také není jasné, z jakého důvodu byla použita právě tato skupina shlukovacích metod.
- Modul IS v příloze 2 zcela neodpovídá textu na straně 25, nebo „cestovní výkazy“ odpovídají „projektům“?
- Kromě základních statistik by měly být v kapitole 6 uvedeny také korelace mezi vstupními atributy.
- V kapitole 6.4.5 postrádám graf Silhouetovy míry.
- Porovnání metod v kapitole 6.4.6 je založeno pouze na počtu shluků. To je dosti zjednodušené i vzhledem k tomu, že autorka pro kvalitu shlukování používala Silhouetovou míru.

- Algoritmus pro segmentaci nových zákazníků není z textu v kapitole 7 zřejmý, je však uveden na příloženém CD.

Formální připomínky:

- V textu je nejprve uveden odkaz na přílohu 2, až pak na přílohu 1.
- Proměnné nejsou zapisovány jednotně, jednou kurzívou, jindy bez kurzívy.
- Zápis MNČ na str. 36 není matematicky přesný.
- V práci je několik pravopisných chyb a překlepů (např. „kohenze“ v kapitole 6).
- Seznam použité literatury není seřazen ani abecedně, ani podle výskytu v textu.

Otázky k obhajobě práce:

1. Vysvětlíte pojem CF-strom a jeho parametry.
2. Jaký algoritmus a s jakým nastavením byl použit pro učení Kohonenovy samoorganizující se mapy?
3. Vysvětlíte vyjádření na str. 46 „záznamy byly zařazeny do skupin intuitivně ... aby jich ve skupině A bylo nejméně“. Podobná vyjádření naznačují, že výsledné hodnocení může být zkreslené použitou metodikou.
4. Jakým způsobem byly určeny hodnoty ukazatelů zisku pro výpočet skóre podle vzorce (10)? Dále vysvětlíte, jakým způsobem bylo určeno maximální skóre 709 bodů. Nemělo by to být spíše 790?

**Práci doporučuji k obhajobě**

**Hodnocení práce známkou:      **Velmi dobře****

V Pardubicích, dne 20. května 2015

doc. Ing. Petr Hájek, Ph.D.

