

Posudek oponenta diplomové práce

Jméno studenta	Bc. Michael Siegert
Téma práce	Analýza dat pro CRM v organizaci
Cíl práce	Realizovat analýzu dat o zákaznících a jejich segmentace pro potřeby CRM. S využitím poskytnutých dat z organizace realizovat jejich analýzu, nalézt charakteristické rysy jejich chování a ukázat na možnosti využití v rámci CRM.
Vedoucí diplomové práce	doc. Ing. P. Petr, Ph.D.

náročnost tématu na	úroveň		
	nadprůměrná	průměrná	podprůměrná
teoretické znalosti		x	
praktické zkušenosti	x		
podkladové materiály (vstupní data) a jejich zpracování	x		

kritéria hodnocení práce	úroveň			
	nadprůměrná	průměrná	podprůměrná	nelze hodnotit
stupeň splnění cíle práce		x		
samostatnost při zpracování tématu				x
logická stavba práce		x		
práce s českou literaturou včetně citací		x		
práce se zahraniční literaturou včetně citací		x		
adekvátnost použitých metod		x		
hloubka provedené analýzy		x		
stupeň realizovatelnosti řešení		x		
formální úprava práce (text, grafy, tabulky)		x		
stylistická úroveň		x		
nároky DP na podkladové materiály, konzultace, průzkumy ...	vysoké x	průměrné	nižší	nejsou
použití analýz, matem. statistických a jiných metod, komparací apod.	ve velké míře	přiměřené x	částečné	absentuje
využitelnost námětů, návrhů a doporučení k řešení problému	ve větší míře x	částečná	nižší	nevyužitelnost
obsah a relevantnost příloh v textu či příl. části DP (tabulky, grafy, propočty apod.)	vysoce funkční	funkční	méně funkční x	neuspokojivé

Připomínky a poznámky k diplomové práci:

Zpracování předložené diplomové práce vychází z metodiky CRISP-DM. Autor pracuje s reálnými daty o klientech poskytnutými nejmenovanou bankovní společností. V úvodní části práce jsou definovány jednotlivé fáze této metodiky a pojem CRM. V následujících (kapitoly 2 – 4) je pak uvedeno porozumění problému včetně definice obchodních a data miningových cílů, porozumění datům a jejich příprava včetně tvorby odvozených proměnných. Modelování je obsahem kapitoly páté, kde se autor zaměřuje na tvorbu modelů s využitím metod K-Means a C5.0. V této kapitole se také zabývá analýzou časových řad. Vyhodnocení výsledků je obsahem kapitoly šesté. Využití dosažených výsledků v praxi je autorem velmi stručně nastíněno v kapitole sedmé.

Díky následování fází uvedené metodiky je práce přehledná. V práci se místy objevují překlepy (např. str. 32, 45, 52, 59), obsahové nepřesnosti a chyby (např. str. 18 v části 3.2 Datový slovník – atribut (proměnná) DRUH OSOBY (s významem *pohlaví a druh osoby* –

v příloze A, kde je uvedeno pouze Pohlaví, a to se 4 kategoriemi (F/M/PODNIKATEL/PRAVNICKA_OSOBA); chybí vysvětlení hodnot atributu N_Klient_Kateg_Obsl); obrázek 2 Data audit na str. 18 neukazuje na data uvedená v datovém slovníku (jedná se již o zobrazení dat včetně odvozených atributů); nepřesnosti ve vzorci (2) na str. 28, na str. 33 – obrázek 7 Vstupy do systému (využití atributu Pohlaví a Klient_Typ); místy nedokončené věty a souvětí a jejich srozumitelnost (např. na str. 17 v kapitole 3. Porozumění datům, na str. 29 v kapitole 4.1.2 Binarizace atd.).

Některé části práce by bylo vhodné více propracovat či doplnit (týká se např. kapitol 3 a 4, které se zabývají porozumění datům a jejich přípravou (např. identifikace odlehlých hodnot; vzájemné vztahy mezi proměnnými; doplnění popisu nově vytvořených proměnných; chybí informace, s jakými proměnnými a s kolika záznamy bylo pracováno po provedených úpravách původních dat atd.); dále kapitoly 5. Modelování (zejména části zabývající se shlukovou analýzou (str. 31) a tvorby modelu pomocí metody C5.0 (str. 37) – autor uvádí vždy využití pouze 1 metody na stávajících datech).

I přes uvedené připomínky je předložená diplomová práce zajímavá, splňuje požadavky na diplomovou práci, a proto ji **doporučuji k obhajobě**.

Navržený klasifikační stupeň:

Velmi dobře

Otázky k obhajobě:

1. Na str. 33 uvádíte: *Pro spojité proměnné modeler provedl standardizaci a normalizaci, u kategorizovaných proměnných binarizaci*. Na základě jakého vzorce dochází k transformaci číselných proměnných (typ Continuous) u metody K-Means realizované v IBM SPSS Modeler do intervalu $<0, 1>$?
2. V kapitole 5 Modelování byla uvedena pouze metoda K-Means s určením 6 shluků, přestože, jak v práci uvádíte, jste provedl shlukování do různého počtu shluků (4 až 9). Na základě čeho byl určen stávající počet shluků jako vhodný, přestože sledované kritérium siluety nevykazovalo příznivé výsledky?
3. V rámci kapitoly 5 Modelování byl použit pouze algoritmus C5.0. Jaké další algoritmy pro tvorbu rozhodovacích stromů by bylo vhodné v dané oblasti použít?
4. Proč jste se rozhodl pro modelování trendu pomocí klouzavých průměrů a na základě čeho jste určil index determinace?

Oponent diplomové práce:

Ing. Miloslava Kašparová, Ph.D.

V Pardubicích dne: 31. 5. 2017