



## POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

**Jméno studenta:** Norbert Snášel

**Název práce:** Data mining: Analýza klientů retailového bankovníctví

**Autor posudku:** prof. RNDr. Hana Skalská, CSc.

**Cíl práce:**

Pomocí vybraných nástrojů analyzovat data, týkající se využívání služeb běžných účtů bank v ČR.

Povinná kritéria hodnocení práce	Stupeň hodnocení (známka)			
	A	C	E	F
Práce svým zaměřením odpovídá studovanému oboru	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vymezení cíle a jeho naplnění	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Zpracování teoretických aspektů tématu	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Zpracování praktických aspektů tématu	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adekvátnost použitých metod, způsob jejich použití	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hloubka a správnost provedené analýzy	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Práce s literaturou	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Logická stavba a členění práce	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Jazyková a terminologická úroveň	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Formální úprava a náležitosti práce	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vlastní přínos studenta	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Využitelnost výsledků práce v teorii (v praxi)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

### Dílčí připomínky a náměty:

Cíle práce by bylo vhodné strukturovat a záměr práce podrobněji vysvětlit v kontextu na data předpokládané aplikace. Logická stavba práce není v metodice vysvětlená a je nevyvážená. Obecné poznatky (například kapitola 3.2 týkající se metodiky CRISP-DM) nepatří do kapitoly 3 (Popis a příprava dat), ale jsou součástí metodologie data mining. Texty v některých kapitolách a často ani ve stejném odstavci nejsou navzájem provázány a přecházejí na vysvětlování doplňků, které náleží do obecnější kapitoly (týká se například vysvětlování možností softwaru, nebo komentářů k výsledkům). Chybí popis výchozích atributů zdrojových dat. Kapitola „Příprava dat“ sice popisuje ošetření jejich nekonzistencí, ale celkový přehled o struktuře výchozích dat chybí.

Popis výsledků práce je nepřehledný. V kapitole 6.2 není uvedena metoda analýzy. Z výstupu plyne, že se jedná o diskriminační analýzu, hypotézy jsou možná uvedeny v jiné části práce. Členění ukázky (aplikace) do dvou kapitol 5 a 6 není vysvětleno a takto není srozumitelné.

Název softwaru je v celé práci uváděn nepřesně. Popisy využitých metod jsou často také velmi zjednodušené až nepřesné. Například obecným výrazem „kvalitněji“ je subjektivně zhodnocena nebo zdůvodněna určitá volba metod nebo postupu (strany: 1 - anotace, 2, 6, 9, 15, 16, 17, 19, 20, 30, 37, 50).

Kvalita modelu nebo postupu řešení by měla být hodnocená objektivně (odborně) a mělo by být vysvětleno a zdůvodněno, jaká kritéria jsou použita a proč (v čem) je navržený postup vhodnější. Data z kalkulátoru bankovních poplatků mohou být fiktivní již od zdroje, zvolené analýzy a především jejich závěry by měly tento aspekt přinejmenším zmínit.

V práci je řada stylistických i terminologických prohřešků a samotný název práce je zavádějící („Analýza klientů...“) Jedná se o analýzu dat (ne klientů). Práce pomíjí nemožnost prokázat, zda uživatel kalkulátoru bankovních poplatků (zdroj dat) je skutečně klientem banky, kterou případně uvedl a zda veškeré uvedené informace (tedy i data na nich založená) jsou validní.

V závěrech práce chybí přehledné celkové shrnutí práce a zhodnocení vytčeného cíle. Závěr popisuje převážně výsledky kapitol 5 a 6.

V práci jsou četné prohřešky stylistické, formální ale též věcné, zvýrazněny jsou v tištěné verzi práce. Příklady nepřesných až nesprávných slovních spojení: ...“pokles v logaritmické pravděpodobnosti“ (str.19), ... „proměnné mají odlišné míry“ (str.31), ...„variace závisle proměnné“ (str. 47). Příklad chyby v aplikaci: Na str. 35 uvedeno nastavení *within group covariance matrix*, na straně 45 zamítnutá hypotéza o shodě kovariančních matic, rozpor by zasloužil aspoň vysvětlení nebo komentář.

### **Celkové posouzení práce a zdůvodnění výsledné známky:**

Práce se zabývá analýzou reálných dat, jejichž věrohodnost je neurčitá. Hlavní přínos studenta vidím ve výsledcích kapitol 3.3 až 3.5 (Příprava dat, Explorace dat a Rady pro úpravu kalkulátoru), které se věnují popisu kroků, vedoucích k odstranění logických chyb v datech a přípravě dat pro analýzu. Metodika práce, teoretická a analytická východiska aplikační části práce, ani popis vlastních výsledků přípravy dat a jejich analýz, nejsou rozčleněny do srozumitelných, jednotlicích a logicky uspořádaných rámců.

### **Otázky k obhajobě:**

1. Jaký je význam metod křížové validace? Které metody používá IBM SPSS Modeler?
2. Jaká vidíte rizika při použití Vámi sestavených modelů?

**Práci doporučuji k obhajobě.**

**Navržená výsledná známka: D - velmi dobře-dobře**

**V Hradci Králové, dne 12. května 2016**

---

podpis