



### Rozsah práce

Práce se rozkládá na 104 číslovaných stranách, z toho samotná textová část na 88 stranách. Jedná se o standardní rozsah odpovídající diplomovým pracím.

### Volba tématu

Téma práce je originální, na katedře demografie a geodemografie PŘF UK v Praze doposud nezpracované. Zvolené téma si zvolil autor sám s ohledem na jeho zkušenosti s touto problematikou, které získal při předchozím studiu na jiné vysoké škole.

### Cíl práce a jeho naplnění

Cílem předkládané práce bylo popsat a následně demonstrovat na modelové úloze principy procesu dolování znalostí z databází, často označovaného jako data mining (DM). Motivace autora k výběru tématu bylo studium demografie a získání větších znalostí o charakteru dat, s kterými demografie pracuje, stejně jako skutečnost, že neexistuje v české odborné literatuře mnoho zmínek o využití těchto metod ve spojení s demografií. Další motivací autora bylo zvýšit povědomí o DM metodách mezi lidmi, kteří se zabývají různorodými demografickými problémy k využití při stanovování určitých zajímavých hypotéz v rámci svých vlastních výzkumů. S ohledem na určitou tematickou novost této práce se domnívám, že vyřčené motivační prvky byly naplněny.

### Struktura práce

Práce má standardní strukturu, text je rozdělen do pěti kapitol včetně *Úvodu* a *Závěru*. Úvodní kapitola seznamuje právě s myšlenkou, proč zpracovat toto téma a poskytuje základní přehled odborné literatury. Druhá kapitola poměrně podrobně popisuje DM, a to jak základní vymezení, tak metody a techniky, stejně jako kriticky rozebírá dostupné softwarové nástroje, které umožňují zpracování dat touto metodou. Třetí kapitola je jakýmsi krátkým, ale pro potřeby práce potřebným, exkurzem do teorií výběrových šetření, včetně nastínění vybraných příkladů statistických šetření, prováděných na našem území. Nejzajímavější kapitolou, která je metodologicko-analytického charakteru, je kapitola čtvrtá, kde autor seznamuje krok po kroku s postupem práce při zpracovávání DM úlohy dle vybrané metodiky (CRISP-DM). Závěr práce shrnuje získané a předložené poznatky a nastiňuje, kam by dané téma mohlo dále směřovat.

### Práce s literaturou

Diskuse s literaturou s ohledem na zvolené téma je v této práci poměrně dostačující, a to i díky skutečnosti, že se jedná o poměrně novou, prozatím ne příliš popsanou metodu. Citace použitých zdrojů je až na malé nepřenosnosti, plynoucí z nepozornosti, správná.

### Metodologické postupy

Oceňuji snahu autora popsat jednotlivé metody, techniky a postupy tak, aby byly co nejvíce srozumitelné. Většinou se to v práci daří, nicméně někdy autor až příliš předpokládá, že je čtenář již „vžit“ se studovanou problematikou, bližší objasnění by však někdy bylo na místě.

### Přínosy práce

Největší přínos vidím v originalitě tématu na půdě domovské katedry a nastínění využití DM v demografii. Domnívám se, že i když nebyly použity ne příliš vhodná data, tak práce ukázala, že je možné tyto metody a techniky využít např. pro klasifikační úlohy za účelem odhadu chybějících údajů (což by bylo možné využít při analýze neúplných demografických dat, ať už z klasických výběrových šetření, tak např. vybraných

demografických údajů v rozvojových zemích). Co se týče hledání nugetů (zajímavých znalostí ve vybraných datech pomocí asociačních pravidel), tak tam si nejsem příliš jist praktického využití v demografii, ale například některé demografy, především ekonomicky či sociologicky zaměřené, může tato možnost použití oslovit.

#### Zásadní připomínky k práci

Domnívám se, že pro práci by bylo vhodnější, kdyby byla použita data za české sčítání lidu – byť byla tato možnost se studentem konzultována a nabídnuta pomoc s jejich zajištěním, student této nabídky nevyužil. I proto mě překvapuje odůvodnění nepoužití těchto dat (na s. 56–57). Také výběr některých charakteristik z American Community Survey pro jednotlivé modelové výpočty nejsou občas vhodně zvolené s ohledem na snahu poukázání na využití v demografii.

#### Grafické prvky a přílohy

V textu je použito několik tabulek, které však nejsou příliš dobře graficky zpracované, navíc zápisy zdrojů dat a poznámek také nejsou příliš zdařilé. Umístění grafů v tabulkách 8–10 bych nedoporučoval, tabulky i grafy mohly být zpracovány jinak. Přílohovou část práce neobsahuje, byť např. tabulky č. 8 a 9 mohly být v této části umístěny, místo v textu.

#### Formální stránka

Z vizuální stránky práce je vidět, že autor práci dokončoval v určitém časovém presu a ne zcela se seznámil s doporučenými pravidly pro zpracování závěrečných prací na naší katedře. Často se vyskytuje nelogické zalomení textu (např. začátky kapitol na konci stránky, zdroje dat k tabulkám na začátku následující stránky apod.). I když mám k této oblasti nemalé výhrady, nedomnívám se, že se jedná o zásadní problém.

#### Jazyková stránka

Po stylistické stránce je práce poměrně dobře sepsaná. Bohužel, v práci zůstaly některé gramatické chyby (např. „byli“ na s. 56 nebo s. 87) či hovorové výrazy (s. 78 „nula letí“).

#### Další hodnocení

Oceňuji snahu studenta připojit svůj názor, např. při hodnocení možného použití výpočtových algoritmů, kde jsou často popsány klady i zápory jednotlivých algoritmů. Co se týče spolupráce studenta v průběhu zpracování práce, tak ačkoliv z počátku to vypadalo na zajímavou spolupráci, nakonec student práci dokončil bez zásadních konzultací a postavil školitele do situace, kdy nebylo možné již práci důkladně pročíst a upozornit na některé z výše uvedených nesrovnalostí.

I přesto se však domnívám, že výsledná práce Ing. Davida Fišera je zajímavá a pro mnohé další možné následovníky v této oblasti přínosná a že práce splňuje požadavky kladené na zpracování závěrečných prací podobného typu, a proto

**navrhuji přijmout překládanou diplomovou práci k obhajobě.**

RNDr. Luděk Šídlo, Ph.D.

Praha 2015-09-03