

Posudek oponenta diplomové práce

1. Identifikační údaje

Název práce: **Návrh predikčního modelu pro odhalování odcházejících zaměstnanců**
Autor práce: **Michaela Vojáčková**
Vedoucí práce: **doc. Ing. Pavel Petr, Ph.D.**

2. Cíl práce

Cílem předložené diplomové práce byl návrh predikčního modelu pro odhalování odcházejících zaměstnanců, jehož naplnění spočívalo v popisu základních pojmů z oblasti predikčního modelování, návrhu a tvorbě predikčního modelu a jeho verifikace. Na základě obsahu předložené diplomové práce lze konstatovat, že autorka splnila požadavky kladené zadáním.

2. Obsahové zpracování práce

Autorka svou práci rozdělila pouze do dvou hlavních kapitol, z nichž druhá kapitola představuje těžiště vlastní práce. Cíl práce je definován v úvodní části práce a dále rozpracován v kapitole druhé. Autorka ve své práci pracuje s „ostrými“ daty nejmenované firmy a postupuje podle metodiky CRISP-DM, která je stručně popsána v kapitole první. Při návrhu predikčních modelů se zaměřuje na vybrané algoritmy rozhodovacích stromů, neuronové sítě a regresní analýzu. Ve fázi předzpracování dat mimo jiné využívá také metody segmentace. V rámci daného řešení využila programové prostředí SPSS Clementine 11.1, SPSS 16.0, MS Excel a ve fázi přípravy a předzpracování dat i ArcGIS 9. Autorka si pro svou diplomovou práci zvolila zajímavou, v současné době velmi aktuální oblast personalistiky. Svou orientaci v řešené problematice prokázala správným výběrem metod a jejich aplikací při návrhu predikčních modelů.

3. Formální náležitosti a úprava

Předložená diplomová práce o rozsahu 70 stran textu je zpracována srozumitelně. Je doplněna 33 tabulkami, 13 obrázky, 23 grafy a 3 přílohami. Z formálního hlediska lze v práci nalézt drobné nedostatky (například str. 7, 33, 43).

4. Připomínky

Z hlediska obsahového zpracování lze práci vytknout:

- Nepřehledné strukturování textu (práce je rozdělena pouze do dvou kapitol) a poměrně malý rozsah použité literatury.

- Pouze stručný slovní popis použitých metod (vhodné je použít metody matematicky popsat).
- Závěrečné vyhodnocení výsledků algoritmů ze str. 39, kapitola 2.3.5 (není vhodné porovnávat výsledky u algoritmů rozhodovacích stromů, které pracovaly na základě různých vstupních proměnných).

Z hlediska interpretace by bylo vhodné převést získaná pravidla (např. obrázky 5, 8, 9) z rozhodovacích stromů na pravidla s původními daty.

K práci nemám jiné zásadní obsahové připomínky.

5. Otázky doporučené k bližšímu vysvětlení


- a) Jaká se používají štěpící kritéria u algoritmů rozhodovacích stromů použitých ve Vaší diplomové práci?
- b) Podle čeho byla na str. 25 stanovena hranice 3 % pro určení odlehlých hodnot a hranice 5 % pro určení extrémů? Jak lze ve Vaší práci chápat pojem „extrém“ a „odlehlá hodnota“?
- c) Blíže vysvětlete graf 6 na str. 35 (část testovací).
- d) Vysvětlete text ze str. 41: „*Rozhodovací strom C5.0: ... na výstupu jedna nebo více symbolických proměnných.*“ Je tento text pravdivý?
- e) Na základě čeho jste se rozhodla rozdělit soubor dat (str. 42) právě na 45 % dat pro trénování, 40 % dat pro testování a 15 % dat pro validaci. Je možné Vaše rozhodnutí podložit nějakou literaturou?

I přes uvedené připomínky předloženou diplomovou práci

doporučuji k obhajobě

a hodnotím ji známkou **velmi dobře**.

Pardubice, 27. května 2009


Ing. Milošlava Kašparová, Ph.D.
Oponent diplomové práce