Posudek bakalářské práce – vedoucí

Autor BP: ***Jakub Dolejšek***
Název BP: ***Získávání znalostí z databází***

Autor práce měl prostudovat známé metody dobývání znalostí z databází (*data mining*) a pokusit se navrhnout nějaké rozšíření „běžných“ statistik nad (reálnými) daty. Bylo doporučeno, aby měla práce i „praktický“ přesah, tj. aby byly vybrané metody aplikovány na skutečných datech.

Práce obsahuje úvod do *data miningu* (DM) a zaměřuje se na úlohy, které DM umí řešit, tj. dozvídáme se, jaké jsou cíle (podrobné) analýzy dat. Centrálním tématem teoretické části práce je seznámení s metodikou analýzy dat (CRISP-DM) a s jejími dvěma důležitými fázemi: (a) porozumění a příprava dat a (b) modelování analyzovaných dat (analytické algoritmy). V (a) byla využita konkrétní data (výsledky UNIFITTESTu) se základní statistikou ukazující některé vztahy mezi výsledky. Zde autor použil grafický výstup pro snadnou orientaci ve výsledcích. V (b) nás autor seznamuje s metodami analýzy dat a s problémem přeučení. Z metod si nakonec vybírá *rozhodovací stromy (algoritmus ID3)* a *neuronové sítě*. Tyto metody jsou popsány podrobněji a autor je aplikuje na data získaná z UNIFITTESTu.

Shrnutí

Autor nastudoval množství teorie kolem DM tak, aby si mohl dvě metody vybrat a ty podrobněji zpracovat. Zpracování proběhlo na reálných datech a vyžadovalo pracnou přípravu softwaru pro obě zmiňované fáze (a,b) – viz rozsáhlá příloha práce. Autor tedy musel připravit poskytnutá data a provést jejich základní analýzu (a). Na tomto základě zvolil dvě metody používané v DM (b) a ty aplikoval a následně i porovnal. Zadání práce tak bylo splněno. Je nutno pochválit i jazykovou úroveň práce.

Práce splňuje náležitosti podle studijního řádu a lze ji doporučit k obhajobě.
Navrhuji známku **výborně**.

V Praze dne 8.9.2013

Michal Peliš
oddělení logiky
Filosofický ústav AV ČR