

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Student: Bc. Pavel Kučera

Oponent: doc. Ing. Pavel Krömer, Ph.D.

Studijní program: Inženýrská informatika

Studijní obor: Informační technologie

Akademický rok: 2014/2015

Téma diplomové práce: Analýza výrobních a zkušebních dat pomocí data miningu

Hodnocení práce:

Předložená práce se zabývá analýzou výrobních dat obchodní společnosti pomocí metod získávání znalostí a dolování z dat (KDDM). Předložený text je rozdělen do tří částí. V první, teoretické, seznamuje autor čtenáře s popisem procesu dolování z dat a jeho jednotlivých etap. Věnuje se zejména oblasti předzpracování dat a popisu některých vybraných metod KDDM. Ve druhé části, nazvané analytická, diplomant popisuje zdroje, strukturu, a sémantiku dat, se kterými se nadále pracuje. Jsou také představeny některé obvyklé nástroje pro KDDM. V poslední, projektové, části je popsána vlastní aplikace procesu KDDM na získaná data.

Předložená práce se zabývá aktuální problematikou a svým rozsahem je dostačující. Formálně splňuje vše, co má kvalifikační práce obsahovat. Její technické zpracování, přesnost uváděných údajů a jazyková stránka však nejsou na nejlepší úrovni. Zejména lze vytknout:

1. nepřesnost/neúplnost – (např. taxonomie na obr. 1.2, tvrzení, že velikost WWW je ‘přes 650 milionů stránek’ atd.)
2. práci se zdroji – autor se často uchyluje k rozsáhlým popisům, definicím a autoritativním tvrzením, která jsou ozdrojována pouze několika referencemi uvedenými na konci oddílu textu (např. sekce 1) nebo vůbec (např. sekce 1.1.3 – 1.1.9). Lze také polemizovat s tím, zda je např. wikipedie vhodným autoritativním zdrojem informací pro odborný text (např. [27]-[30])
3. neaktuální volbu popsanych a použitých algoritmů – autor vychází při identifikaci ‘nejrozšířenějších KDDM algoritmů’ z publikace z r. 2006
4. zaměření na technické detaily – např. popisy nástrojů místo algoritmů v tab. 2.1
5. značnou neformálnost v projektové části. Data by měla být popsána pomocí nějakého, byť třeba odpovídající způsobem zjednodušeného, formalismu (E-R model, UML model). Přiložené kusovníky a výpisy kódu považuji v textu práce za nadbytečné.
6. slabou jazykovou kvalitu – předložený text také obsahuje značné množství překlepů a elementárních pravopisných a gramatických chyb (např. v závěru)

Uživatelská studie, aplikující vybrané algoritmy na reálná provozní data, je zajímavá, vyhodnocení výsledků je však vágní (sekce 7.4). Kupříkladu není příliš snadno uchopitelné autorovo konstatování ‘Z analýzy shluků a zároveň také z logistické regrese a neuronové sítě vyplývá, že pro dosažení vyšší pravděpodobnosti výsledku "neshodný výrobek" je třeba nastavení několika parametrů zároveň.’. Jak to vyplývá? Jaké parametry je třeba nastavit? Diplomant by toto měl vyjasnit při obhajobě.

I přes uvedené připomínky se domnívám, že práci lze doporučit k obhajobě, ale nezbyvá mi, než ji hodnotit pouze jako dobrou.

Celkové hodnocení práce:

Známku uvede oponent dle svého uvážení dle klasifikační stupnice ECTS:

A – výborně, B – velmi dobře, C – dobře, D – uspokojivě, E – dostatečně, F – nedostatečně.

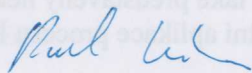
Stupeň F znamená též „nedoporučuji práci k obhajobě“.

Předloženou diplomovou práci doporučuji k obhajobě a navrhuji hodnocení

C - dobře.

V případě hodnocení stupněm „F – nedostatečně“ uveďte do připomínek a slovního vyjádření hlavní nedostatky práce a důvody tohoto hodnocení.

Datum 21. 5. 15


Podpis oponenta diplomové práce