

Review of Master's Thesis

Student: Sabovčík František, Bc.
Title: Data Mining in Small Business (id 20931)
Reviewer: Burgetová Ivana, Ing., Ph.D., UIFS FIT VUT

1. **Assignment complexity** **average assignment**
2. **Completeness of assignment requirements** **assignment fulfilled**

Zadání bylo splněno v plném rozsahu. Za určité rozšíření považuji využití hned několika modelů pro predikci hodnot časových řad, hledání vhodných parametrů některých z těchto modelů pomocí genetických algoritmů a využití několika metrik pro hodnocení kvality získaných výsledků.
3. **Length of technical report** **in usual extent**
4. **Presentation level of technical report** **62 p. (D)**

Prezentační úroveň této práce není příliš dobrá. Zpráva je špatně srozumitelná, málo konkrétní a nejasná. (Např. v kapitole 3 není jasné, proč jsou popisovány zrovna uvedené 3 softwarové nástroje, z kapitoly 4 není příliš jasné, kde se berou data, která student dále analyzuje, kapitola 5 není nijak strukturovaná, chybí vzorec 7.6 atd.) Stejně tak struktura práce není ideální. Není zde dobře oddělen teoretický úvod od vlastních analýz (kapitola 8 popisuje použité modely, jejich implementaci i dosažené výsledky).
5. **Formal aspects of technical report** **63 p. (D)**

Jazyková stránka předložené práce není dobrá. Zpráva obsahuje řadu špatných jazykových konstrukcí, což velmi stěžuje pochopitelnost celé zprávy. Co se týká typografie, je zřejmé, že student se snažil vytvořit kvalitní práci. Nicméně ve zprávě zůstalo poměrně dost nekonzistencí (str. 28 - nedokončený odrážkový seznam, strana 6 - sazba proměnných normálním písmem, str. 32 - neoddělení definic od textu, atd.)
6. **Literature usage** **60 p. (D)**

Seznam použité literatury je značně obsáhlý a obsahuje povětšinou kvalitní relevantní zdroje (bohužel některé citace nejsou úplné, např. chybí ISBN). Poněkud horší je to s uváděním zdrojů v samotné práci. Např. informace v kapitolách 8.4 a 8.8 byly zřejmě odněkud převzaty, ale zdroj není uveden. Na druhou stranu nelze říci, že by se student snažil porušit citační etiku.
7. **Implementation results** **70 p. (C)**

Výstupem této diplomové práce je program v pythonu, který umožňuje pro zadaná data vytvořit vybraný model pro predikci hodnot časových řad. Dále pak tento model otestovat na dostupných datech a ohodnotit jeho kvalitu pomocí různých metrik. Pro predikci jsou k dispozici 4 modely. Dále pak student vytvořil program v pythonu pro dolování frekventovaných množin, který ale ve zprávě není zdokumentován. Vytvořené zdrojové kódy jsou funkční, ale nejsou bohužel téměř vůbec zdokumentované.
Student vytvořené programy využíval na analýzu dat malých podniků a získané výsledky jsou součástí technické zprávy.
8. **Utilizability of results**

Za přínos této práce považuji otestování vlastností 4 různých modelů pro predikci hodnot časových řad (predikce tržeb malých podniků).
9. **Questions for defence**
 - Který z vytvořených modelů dosahoval nejlepších výsledků? Lze doporučit jeden konkrétní model pro všechny testované podniky, nebo pro různé podniky jsou nejvhodnější různé modely?
10. **Total assessment** **67 p. satisfactory (D)**

Jedná se o poněkud rozporuplnou diplomovou práci. Realizační výstup je funkční a technická zpráva obsahuje řadu kvalitních informací i cenných výsledků. Bohužel tyto informace nejsou příliš dobře strukturovány a nejsou prezentovány dobrou formou. Proto navrhuji hodnocení stupněm D.

In Brno 7. June 2018

.....
signature

