

Posudek vedoucího diplomové práce

Diplomová práce studenta **Jana Jakeše** se zabývá implementací metod pro vyhodnocování metod shlukové analýzy pro nečíselná data. K předložené práci mám následující výhrady a komentáře:

- V práci se mi líbí řešerše shlukovacích metod a metrik použitelných pro numerická, nominální a mixovaná data. Stejně tak i popis metod pro vyhodnocení výsledků shlukování v kapitole 4. Zde by nicméně neuškodily obrázky nebo příklady výpočtu.
- Student se dokázal zorientovat poměrně rozsáhlém a ne dobře dokumentovaném projektu Rapidminer a rozšířit jej o požadovanou funkcionalitu.
- Popis SW Rapidminer by si zasloužil více textu než jen jeden odstavec, který tvoří celou kapitolu 5. Tato kapitola by buď měla obsahovat více informací nebo se měla stát součástí kapitoly 6, kde student popisuje implementaci rozšíření Rapidmineru.
- V kapitole 6 bych očekával alespoň krátký popis API na straně Rapidmineru, které mohou jako autor rozšíření využít.
- V kapitole 7 kladně hodnotím snahu studenta srovnat výsledky jeho implementace s implementací v jiných nástrojích.
- U implementovaných operátorů uvádíte čas potřebný pro výpočet výsledků, na jaké HW konfiguraci jste výsledky získal? Máte představu proč, na relativně malých datech — cca 20 tisíc záznamů, výpočet výsledků jednotlivých operátorů trvají jednotky i desítky minut (např pro operátor Silhouette i 25 minut)? Jaká je asymptotická složitost výpočtu jednotlivých operátorů? Jak ovlivní čas výpočtu počet vstupních atributů?
- Kolik dalšího úsilí by bylo třeba, aby bylo možné umístit vaše operátory na Rapidminer Marketplace?

Celkově hodnotím předloženou práci pozitivně - student se dobře zorientoval v problematice shlukovacích metod a jejich vyhodnocování. Dále musím vyzdvihnout fakt, že se student dokázal zorientovat v relativně rozsáhlém a špatně dokumenovaném kódu SW Rapidminer a doplnit požadovanou funkcionalitu. Jen škoda, že nezbylo více času na důkladnější testování na více různých datech, učesání textu práce a případné umístění kódu na RapidMiner Marketplace.

Konstatuji student splnil všechny body zadání. Proto práci doporučuji k obhajobě a hodnotím stupněm

B - velmi dobře

Miroslav Čepek