

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Clustering Social Network Texts
Jméno autora:	Egor Sarana
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra kybernetiky
Vedoucí práce:	Ing. Jan Drchal, PhD.
Pracoviště vedoucího práce:	AIC

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadání obnáší jak sběr a předzpracování dat, tak testování vektorových reprezentací a shlukovacích algoritmů.	

Splnění zadání	splněno s menšími výhradami
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Zadání považuji za splněné kromě bodu 6, tj. otestování algoritmů na datech z Twitteru nebo Telegramu dodanými školitelem. Vzhledem k nedostatku komunikace (viz níže) a času na konci semestru jsem tato data nemohl dodat. Student přesto testoval na datové sadě postavené na Twitteru, kterou si sám vytvořil a kterou považuji za zdařilou.	

Aktivita a samostatnost při zpracování práce	D - uspokojivě
<i>Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.</i>	
Student se mnou bohužel takřka po celý semestr nekomunikoval (ozval se až 19. prosince), proto jsem neměl šanci práci příliš ovlivnit. Na druhou stranu jej musím pochválit za samostatnost, se kterou si dohledal příslušné metody, vyrobil datovou sadu a navrhl a naimplementoval celou evaluační pipeline. I když práce není úplně dotažená, není špatná.	

Odborná úroveň	C - dobře
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Z odborného hlediska je práce na dobré úrovni, i když se tematicky ubírala trochu jiným směrem, než jsem si jako vedoucí představoval. Plánoval jsem s více datasetů a evaluaci na reálnějších datech založenou na ručních anotacích (ne nutně studentových) a shlukování pro neznámý počet clusterů. Přesto jsem však s výstupem spokojen – největším přínosem je studentem vytvořená datová sada, kterou vytvořil syntézou více datových zdrojů, přičemž chytrě získal označení jednotlivých shluků. Bylo by zajímavé vytvořit více verzí této sady (např. balancovanou verzi, verze různých velikostí, apod.) a porovnat vlastnosti shlukovacích metod obecněji. V experimentální části postrádám porovnání metod ve verzích bez redukce dimenzionality (UMAP a SVD), hlavně ale chybí hypotézy a podrobnější analýza, vysvětlující chování různých metod.	

Formální a jazyková úroveň, rozsah práce	C - dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psána kvalitní, srozumitelnou angličtinou. Rozsah textu je průměrný a až na finální kapitoly dostatečný. Za problematický považuji chybějící závěr a hlavně málo detailní rozbor výsledků. S komentářem „ <i>It is hard to fit all metrics to PDF width, so some less important metrics were pruned...</i> “ se příliš neztotožňuji, stejně jako s prezentací nejdůležitějších výsledků ve formě dvou celostránkových tabulek (ty by se hodily spíše do příloh). Práce obsahuje množství obrázků, velká část je však převzatá a v textu nejsou ilustrace dostatečně vysvětleny (např. obrázek 3.1).	

Výběr zdrojů, korektnost citací

D - uspokojivě

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Počet citací je pro bakalářskou práci v pořádku, přesto v textu chybí reference některých metod (např. CLARA, AMI, sekce 5.5). U mnoha citací chybí publikace, ve které se objevily.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.

Přestože výsledná práce neodpovídá zcela mým původním představám, konstatuji, že objem i kvalita odvedené práce jsou nezanedbatelné.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **D - uspokojivě**.

Datum: 30.1.2024

Podpis: