

OPONENTNÍ POSUDEK ZÁVĚREČNÉ KVALIFIKAČNÍ PRÁCE

Autor závěrečné práce: Tomáš Krechler

Název práce: Analýza textů online periodik pomocí metod strojového učení

Oponent práce Ing. Michal Rott

Pracoviště oponenta HCL Technologies Czech Republic s.r.o.

A. Kvalita abstraktu, klíčová slova odpovídají náplni práce	Velmi dobře (2)
B. Rozsah a zpracování rešerše	Velmi dobře (2)
C. Řešení práce po teoretické stránce	Velmi dobře (2)
D. Vhodnost, přiměřenost použité metodiky	Výborně (1)
E. Úroveň zpracování výsledků a diskuse	Velmi dobře (2)
F. Vlastní přínos k řešené problematice	Velmi dobře mínus (2-)
G. Formulace závěru práce	Velmi dobře mínus (2-)
H. Splnění zadání (cílů) práce	Splněno
I. Skladba, správnost a úplnost citací literárních údajů	Dobře (3)
J. Typografická a jazyková úroveň (vč. pravopisu)	Velmi dobře mínus (2-)
K. Formální náležitosti práce (struktura textu, řazení kapitol, přehlednost ilustrací)	Velmi dobře (2)

Komentáře či připomínky:

Po formální stránce mi nejvíce vadilo vkládání výstřžků obrazovky tam, kde nejsou potřeba (např. textové výstupy s funkce). Nevím, co znamenají podivná písmena v seznamu literatury u blogu A. Radforda [6],[7]. I když je použita Pytorch knihovna, tak je v kapitole "4.2.1 Definice a inicializace modelu" sáhodlouze popisován kód trénování modelu. Je vidět nezkušenost autora ve psaní akademických/vědeckých textů. To je naprosto pochopitelné, jelikož se jedná o jeho první text. Doporučil bych častěji konzultovat práci s vedoucím. Práce by měla být Vaše vizitka a to i její úprava.

V práci mi chybí popis vytvořeného datasetu. Kromě statistických dat by mě zajímal i formát, v jakém je uložen a jak se k datům přistupuje.

Nápad generovat komentáře je velmi zábavný. Bylo by zajímavé zveřejnit takové komentáře a sledovat jejich přijetí u komentujících. Problém s chybami v komentářích bych doporučil řešit pomocí automatické korekce. Častější jev v komentářích je komentátorova chyba než výskyt nového slova.

Autor nevzal v potaz ohebnost českého jazyka, i když na to poukázal. Generoval velmi velké slovníky s velkým podílem unikátních slov (výskyt jen jednou v datasetu). Autor tyto výskyty zahodil, i když měl k dispozici jen malý dataset. Lemmatizace těchto slov je zajímavější přístup.

...pokračuje na straně 2

Celkové zhodnocení:

Práci hodnotím pozitivně. Autor sice píše, že nemá pro obor IT valný význam, ale určitě se naučil využívat zajímavé technologie. Ocenil bych raději rozsáhlejší diskuzi nad výsledky než popis programového kódu.

Autor si také vytyčil cíl, který není uveden v zadání. Chtěl provést klasifikaci sentimentu. Toto je velmi rozsáhlé téma výzkumu. Bohužel v rešeršní části práce se autor nezmiňuje o metodách, které jsou využívány, a věnuje problematice jen 5 řádků. V implementační části je také nedostatečně popsána.

Otázky k obhajobě:

1. Jak je vypočítáno výsledné skóre podobnosti funkcí `most_similar` a `similarity`?
2. Jak je možné pomocí metrik `precision` a `recall` ošálit zákazníka na prodat mu špatný systém jako výborný?

Celková klasifikace:

Práce splňuje požadavky na udělení akademického titulu, a proto ji doporučuji k obhajobě

Navrhuji tuto práci klasifikovat stupněm **Velmi dobře (2)**

V Liberci

dne 13. 1. 2021

Podpisem současně potvrzuji, že nejsem v žádném osobním vztahu k autorovi práce

.....

podpis oponenta