FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Oponent práce:	Ing. Kamil Dedecius, Ph.D.
Student:	Bc. Tomáš Valenta
Název práce:	Analýza nálady v komunikačních aplikacích
Obor / specializace:	Znalostní inženýrství
Vytvořeno dne:	5. června 2023

Hodnotící kritéria

1. Splnění zadání

- ▸ [1] **zadání splněno**
  [2] zadání splněno s menšími výhradami
  [3] zadání splněno s většími výhradami
  [4] zadání nesplněno

Zadání mi přijde jako splněné, byť k míře splněnosti mám zejména v 6. bodu menší výhradu.

2. Písemná část práce 65 / 100 (D)

Diplomová práce čítá celkem 71 arabsky číslovaných stran rozdělených do 6 hlavních kapitol. Struktura je zvolena vhodně, má logickou návaznost. Rozsah jednotlivých kapitol je poněkud nevyvážený. První kapitola (Úvod) je sice na dvou stranách, nicméně ta druhá obsahuje 4 slova poslední věty. Druhá kapitola (Teoretická východiska práce) popisuje možnosti, kterak s tématem naložit. Je poměrně obsažná z hlediska záběru, současně ale poněkud heslovitá. Heslovitosti se autor nevyhnul ani v kapitole následující (Existující řešení) o rozsahu tří stran. Těžiště vlastního přínosu spočívá v kapitole čtvrté (Implementace a výsledky). Ta je relativně rozsáhlejší, pokrývá 22 stran, byť i zde by bylo možné jít trochu více do hloubky. Celkově mi přijde z práce nejpovedenější. Pátá kapitola (Testování) je věnována praktickému testu na konkrétní firmě. Její rozsah je čtyři strany, přičemž už na druhé je závěr a pokračují jen 2 strany grafů. Poslední kapitola je jednostránkový závěr.

Jazyk: Práce je napsána v českém jazyce, nicméně obsahuje značné množství anglických slov (forecasting, flow), občas s počeštěním (Dokument-level, workspacu, multivarietní, requesty). Celkově je provázanost s angličtinou silnější, než bych viděl jako žádoucí, a to jsem osobně odpůrcem nucených překladů terminologie do češtiny. Situaci dobře ilustruje seznam zkratek na římsky číslované 10. straně, či obrázky 2.1 a 2.15 na str. 5 a 26. Viz píšeme bez tečky. Práce obsahuje poměrně velké množství překlepů (konrétně,

vyextahovat, základnú formu, mnohadimenziobálním, pravděpodovnost, Jenkisne, stední, nástoroje, odpoveď...), opakovaně nesprávných tvarů (standartizace, pravděbodobnost - viz kap. 2.2.5.4, kde je tento tvar dominantní) a občasné gramatické chyby (chybějící interpunkci, společnostma, pozorování mohly). Občas se objeví špatná stavba věty (str. 4: "typy byly předpokládaly", tamtéž "k objektech", str. 5: "množina všechn slov", str. 15: "prostor funkcí [...] takové", str. 19: "počet špatně klasi v kolik dat", str. 22: "Dále pak Pokud jsou pak [...]", str. 36: "získá vráceno 3 grafy", str. 42: "skutečné logistická regrese"). Rovněž užití velkých písmen je nejednotné (např. str. 4: "Například phase dependency parser, Conditional Random Field [...]"). Na str. 49 začíná věta "Na první pohled vidíme", avšak dál už nepokračuje.

Sazba: Sazbu hodnotím jako solidní. Relativně často dochází k přetečení délky řádku (str. 1, 4, 13, 15, 21...), občas se objeví osiřelé řádky (str. 2, 12, 22). Sazba obrázků je problematická - velmi často jsou špatně čitelné, nejlépe to dokládá kap. 5. Sazbu rovnic by bylo možné rovněž vylepšit, viz např. def. 2.18, kde čárka patří za rovnici, či def. 2.17 a 2.20 se zvláštním znakem u levé hranaté závorky. Jednopísmenné spojky a předložky by neměly stát na koncích řádků (byť zrovna o tomto pravidle si myslím své). Násobení neznačíme hvězdičkou (str. 7). Na str. 10 (a jiných) chybí interpunkce u rovnice a není přeložena odmocnina. Na str. 24 chybí odřádkování u definice náh. procházky. Odkazy do literatury by měly být součástí věty a stát tedy před tečkou. Na str. 12 je prázdná odrážka, na str. 22 a 25 nikam nepatřící citace. V kap. 4 se opakovaně vyskytují vadné odkazy, jež byly při překladu nahrazeny otazníky "??".

Citace: Autor cituje 37 pramenů. Přehled literatury by zasloužil trochu pozornosti (např. položka 21). Zřejmě ve snaze nedopustiti se prohřešku vůči citační etice autor přebírá celé věty, které překládá do češtiny. Někdy jde dokonce o velmi obecná tvrzení. To indikuje jisté nepochopení principu práce se zdroji. Dobře to ukazuje úvodní odstavec kap. 2.2.2.

Výhradu bych měl následujícím věcem:
- str. 8: "Hodnotu lineárního výrazu, z lineární regrese, pak dosadíme do funkce [...]" - toto je velmi zavádějící popis toho, jak funguje logistická regrese.
- str. 10: "S použitím těchto jednoduchých, leč účinných, metrik [...]" - o účinnosti metriky nemůže být řeč.
- str. 13: suma přes Y' nebude fungovat.
- proměnné nejsou často řádně zavedeny (např. rovnice 2.1)
- str. 18: "měřítko výkonnosti modelu" - modely nejsou výkonné
- str. 24: "Postupně tedy nasčítáváme bílé šumy" - autor měl na mysli jednotlivé veličiny, nikoliv celé procesy.
- str. 25: "[...] dlouhodobý růst st[ř]ední hodnoty" - nejen růst, ale i pokles
- str. 25 a 26: Aditivní a multiplikativní modely obsahují "- t". Není mi jasné proč, ale i kdyby tomu tak bylo, logaritmování na str. 26 nebude fungovat.
- Kap. 4.2.1 - moc mi není jasná transformace do intervalu [-1, 1]. V čem je její smysl? Vždyť logistická regrese pracuje na [0, 1], což by i pro účel práce bylo vyhovující.

3. Nepísemná část, přílohy

95 /100 (A)

Součástí práce jsou zdrojové kódy hostované na gitlabu fakulty. Zdají se být rozumně okomentované. Jako neprogramátor nicméně nedisponuji větší erudicí pro jejich posuzování a doporučuji řídit se především názorem vedoucího práce.

4. Hodnocení výsledků, jejich využitelnost 80 /100 (B)

Diplomová práce se zabývá poměrně zajímavým tématem analýzy a předpovídání nálady v komunikačních aplikacích. Autor dospěl k poměrně jednoduchému a zřejmě rozumně fungujícímu řešení. Osobně se domnívám, že vyvinutý nástroj by mohl mít velmi zajímavý potenciál v oborech souvisejících s globální bezpečností. Např. ruské telegramové kanály zaměřené na válku na Ukrajině poskytují velké množství relevantních dat, jejichž kvalitní analýza by mohla mít značný oborový přesah.

Hodnotím zde 80 body, jelikož z kapitoly 5 - Testování - není možné vyvodit mnoho závěrů.

Celkové hodnocení 75 /100 (C)

Silnou stránkou práce je zejména potenciál mezioborového přesahu. Jako slabinu vidím písemnou část, která je dle mého názoru zatížena větším množstvím nedokonalostí. Práci doporučuji k obhajobě.

Otázky k obhajobě

1. Vysvětlete prosím smysl transformace z oboru [0, 1] do [-1, 1], viz kap. 4.2.1.
2. Bylo by možné přímočaře nasadit vyvinutý nástroj na data v kanálech Telegramu? Jaké problémy lze případně očekávat a jak byste je řešil?

Instrukce

Splnění zadání

Posuďte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

Písemná část práce

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

Nepísemná část, přílohy

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

Hodnocení výsledků, jejich využitelnost

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Celkové hodnocení

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.