



Posudek oponenta závěrečné práce

Student: Bc. Michael Mikuš
Oponent práce: Ing. Magda Friedjungová
Název práce: Utilizing AI/ML methods for measuring data quality
Obor: Znalostní inženýrství

Datum vytvoření: 23. 8. 2020

Hodnotící kritérium:	Způsob hodnocení – následující škálou 1 až 4:
1. Splnění zadání	1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
Popis kritéria: 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.	
Komentář: Zadání práce bylo splněno.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
2. Písemná část práce	80 (B)
Popis kritéria: 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. 26/2017, 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.	
Komentář: Práce je přehledně a logicky členěna. Teoretická část práce je mnohem obsáhlejší než část praktická. Student velmi svědomitě a poněkud zešíroka popsal teorii okolo dat jako takových a plynule přešel k datové kvalitě, které se práce věnuje primárně. Myslím však, že teoretická část mohla být stručnější vzhledem k zadání práce. Student čtenáře dostatečně seznamuje jak se stávajícími metodami pro měření datové kvality, tak s AI/ML alternativami (autoenkodér a dolování asocičních pravidel), které zvolil pro své experimenty. Experimenty samotné student popisuje dostatečně, avšak jejich rozsah by v porovnání s teoretickou částí a se stupněm závěrečné práce mohl být větší. Po jazykové stránce je práce na dobré úrovni, je dobře čitelná a může posloužit jako dobrý teoretický přehled v oblasti datové kvality. Student řádně cituje.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
3. Nepísemná část, přílohy	88 (B)
Popis kritéria: Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využity od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů	
Komentář: Téma práce je experimentálního charakteru. Provedené experimenty student dostatečně popsal ve své závěrečné práci, přiložený kód je řádně zdokumentovaný, strukturovaný a znovuspustitelný. Experimenty byly provedeny na veřejně dostupných datasetech, a tak jsou reprodukovatelné.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
4. Hodnocení výsledků, jejich využitelnost	85 (B)
Popis kritéria: 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.	

Komentář:

Pro měření datové kvality student zvolil dvě AI/ML metody - autoenkodér a dolování asociačních pravidel. Obě metody jsou vyhovující pro řešenou úlohu a zároveň určitým způsobem atypické, což práci dodává výzkumný potenciál. Student zmiňuje několik způsobů, jak lze na jeho práci navázat. Práci vč. zdrojových kódů zpřístupnil veřejnosti na platformě GitHub. Co se týče textu práce, obsáhlá a dobře zpracovaná teoretická část může posloužit nejen jako úvod do problematiky datové kvality, ale i jako ucelený přehled terminologie a souvisejících metod pro její měření.

Hodnotící kritérium:

Způsob hodnocení – nehodnotí se

5. Otázky k obhajobě

Popis kritéria:

Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odřázkami).

Otázky:

K práci nemám žádné otázky.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

6. Celkové hodnocení

85 (B)

Popis kritéria:

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.

Text hodnocení:

Student splnil zadání práce v celém rozsahu. Dostatečným a srozumitelným způsobem popisuje řešenou úlohu včetně provedených experimentů a dosažených výsledků. Na práci lze navázat experimenty, které zváží složitější metody/modely z oblasti AI/ML. Práci navrhuji hodnotit stupněm B a to především z důvodu upřednostnění teorie oproti praxi, na kterou by dle mého názoru měl být kladen větší důraz, neboť v ní má student prezentovat vlastní přínos.

Podpis oponenta práce: